



NUMBER OF SEQ ID NOS: 8252  
 SEQ ID NO 7561  
 LENGTH: 165  
 TYPE: PRT  
 ORGANISM: *Acinetobacter baumannii*  
 US-09-328-352-7561

Query Match 61.5%; Score 496; DB 4; Length 165;  
 Best Local Similarity 61.0%; Pred. No. 4.9e-53;  
 Matches 94; Conservative 25; Mismatches 35; Indels 0; Gaps 0;

QY 2 IRKMGVDYRNDGHIILGKVIPEKGLKHAHSDGVVLAALADALGAAAGDIGH 61  
 DB 10 IFIGQMDVHAEEGNEFVLAGVQIPIHTGLKHAHSDGVVLAALADALGAAAGDIGH 69  
 QY 62 PDDTDPNFKGADSRVLLRHVYGIYKKGKYLVAADVTTIAQPKMLPHVPGKRAIADL 121  
 DB 70 PDDTDPNFKGADSRVLLRHVYGIYKKGKYLVAADVTTIAQPKMLPHVPGKRAIADL 129  
 QY 122 ETDVDPIVTKATTTTEKGFEGKREGIAVQAVYLI 155  
 DB 130 NVDLNDISIKATTTTEKGFEGKREGIAVQAVYLI 163

RESULT 3  
 US-09-198-452A-585  
 Sequence 585, Application US/09198452A  
 Patent No. 6559294  
 GENERAL INFORMATION:  
 APPLICANT: Griffiths, R.  
 TITLE OF INVENTION: *Chlamydia pneumoniae* genomic sequence and polypeptides, fragments thereof and uses thereof, in particular for the diagnosis, prevention and treatment of infection  
 FILE REFERENCE: 9710-003-999  
 CURRENT APPLICATION NUMBER: US/09/198,452A  
 PRIOR FILING DATE: 1998-11-24  
 NUMBER OF SEQ ID NOS: 6849  
 SEQ ID NO 585  
 LENGTH: 199  
 TYPE: PRT  
 ORGANISM: *Chlamydia pneumoniae*  
 US-09-198-452A-585

Query Match 19.0%; Score 153; DB 4; Length 199;  
 Best Local Similarity 29.4%; Pred. No. 9.3e-11;  
 Matches 47; Conservative 26; Mismatches 75; Indels 12; Gaps 5;

QY 3 RYGMGVDFR---NGDHIILGKVIPEKGLKHAHSDGVVLAALADALGAA---ALG 56  
 DB 40 RKGIGDSHRELPESSTKCIIGIIFDCPGFQANSDGDIIFHAIQNAISSVTNKIILG 99  
 QY 57 DIGKHPDDDPNFKG-ADSRVLLRHVYGIYKKGKYLVAADVTTIAQPKMLPHVPGKRA 115  
 DB 100 KVADELQI---RGITDSGIYIEALKSLK-PNKGISHVAITIGSRKFKLSALRQ 154  
 QY 116 NTAADLETVDPIVTKATTTTEKGFEGKREGIAVQAVYLI 155  
 DB 155 NTAQVNLTPDTDIGITATSGEGLSDFGCDGVQCCVLTIV 194

RESULT 4  
 US-09-252-991A-24743  
 Sequence 24743, Application US/09252991A  
 Patent No. 6551795  
 GENERAL INFORMATION:  
 APPLICANT: Marc J. Rubenfield et al.  
 TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO PSEUDOMONAS  
 TITLE OF INVENTION: AERUGINOSA FOR DIAGNOSTICS AND THERAPEUTICS  
 FILE REFERENCE: 107196.136  
 CURRENT APPLICATION NUMBER: US/09/252,991A  
 PRIOR FILING DATE: 1999-02-18  
 PRIOR APPLICATION NUMBER: US 60/074,788  
 PRIOR FILING DATE: 1998-02-18

PRIOR APPLICATION NUMBER: US 60/094,190  
 PRIOR FILING DATE: 1998-07-27  
 NUMBER OF SEQ ID NOS: 33142  
 SEQ ID NO 24743  
 LENGTH: 493  
 TYPE: PRT  
 ORGANISM: *Pseudomonas aeruginosa*  
 US-09-252-991A-24743

Query Match 10.0%; Score 81; DB 4; Length 493;  
 Best Local Similarity 30.9%; Pred. No. 0.26;  
 Matches 29; Conservative 11; Mismatches 32; Indels 22; Gaps 5;

QY 5 GMGVDFRNDGHIILGKVIPEKGLKHAHSDGVVLAALADALGAAAGDIGH 62  
 DB 185 GLGQVH---GDDVVGILALADHEDFRGTDHVDADLAH-----VTLG--GGHV 229  
 QY 63 PDDTDPNFKGADSRVLLRHVYGIYKKGKYLVAADVTTIAQPKMLPHVPGKRAIADL 96  
 DB 230 ----DVAQADDLVHPRHAFGAVGQRGHGLCPAD 258

RESULT 5  
 US-09-252-991A-26787  
 Sequence 26787, Application US/09252991A  
 Patent No. 6551795  
 GENERAL INFORMATION:  
 APPLICANT: Marc J. Rubenfield et al.  
 TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO PSEUDOMONAS  
 TITLE OF INVENTION: AERUGINOSA FOR DIAGNOSTICS AND THERAPEUTICS  
 FILE REFERENCE: 107196.136  
 CURRENT APPLICATION NUMBER: US/09/252,991A  
 PRIOR FILING DATE: 1999-02-18  
 CURRENT FILING DATE: 1998-02-18  
 PRIOR APPLICATION NUMBER: US 60/074,788  
 PRIOR FILING DATE: 1998-07-27  
 NUMBER OF SEQ ID NOS: 33142  
 SEQ ID NO 26787  
 LENGTH: 560  
 TYPE: PRT  
 ORGANISM: *Pseudomonas aeruginosa*  
 US-09-252-991A-26787

Query Match 9.9%; Score 79.5; DB 4; Length 560;  
 Best Local Similarity 23.3%; Pred. No. 0.48;  
 Matches 41; Conservative 34; Mismatches 78; Indels 23; Gaps 7;

QY 1 MIRVMGCV-DVHRND-----GDHIILGKVIPEKGLKHAHSDGVV---LHA 44  
 DB 81 LIRVGLAFRGHLRLADEAHLAALGLVGSANVLDDLGVGRQHLVHLLDQATVGDILQIA 140  
 QY 45 LA--DAIIGAALADIGHPPDDPNFKGADSRVLLRHVYGIYKKGKYLVAADVTTIA 101  
 DB 141 LGDDIVGTLA---VGHGFEDHGLDLAG--DGVVLDADQDHAQLLGGHGLLDQAVIVE 196  
 QY 102 QAPKMLPHVPGKRAIADLETVDPIVTKATTTTEKGFEGKREGIAVQAVYLI 157  
 DB 197 QAAQLDHRHPVRQGLVADALDILEVVGHLAAGQHGAGIVGQAVLALAEAPALLHR 252

RESULT 6  
 US-09-252-991A-32625  
 Sequence 32625, Application US/09252991A  
 Patent No. 6551795  
 GENERAL INFORMATION:  
 APPLICANT: Marc J. Rubenfield et al.  
 TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO PSEUDOMONAS  
 TITLE OF INVENTION: AERUGINOSA FOR DIAGNOSTICS AND THERAPEUTICS  
 FILE REFERENCE: 107196.136  
 CURRENT APPLICATION NUMBER: US/09/252,991A  
 PRIOR FILING DATE: 1999-02-18  
 PRIOR APPLICATION NUMBER: US 60/074,788

; PRIOR FILING DATE: 1998-02-18  
 ; PRIOR APPLICATION NUMBER: US 60/094,190  
 ; PRIOR FILING DATE: 1998-07-27  
 ; NUMBER OF SEQ ID NOS: 33142  
 ; SEQ ID NO 32625  
 ; LENGTH: 562  
 ; TYPE: PR1  
 ; ORGANISM: Pseudomonas aeruginosa  
 ; US-09-252-991A-32625

Query Match 9.7%; Score 78.5; DB 4; Length 562;  
 Best Local Similarity 24.3%; Pred. No. 0.64;  
 Matches 35; Conservative 25; Mismatches 63; Indels 21; Gaps 5;

QY 16 GDHILGGVXIPY--EKGEASHSDGVDVHLALADALIGAAALGDICKHFPDTPDKGAD 73  
 DB 252 GEHVVDPOIRLQITGIGERAGGGGLVTOALFEGGSGRTTIGQ-----TORGAD 301  
 QY 74 SRVLRHYVYGIYKKGKYLVDVTIIAQAAPKMLPHVP--GMRANIADLETDVDFINVKA 132  
 DB 302 ---LLOHVFROPLOQRQALLFEGAEV-----OPTAHPPLGDRGDLADAGLGQFVDDER 353  
 QY 133 TTTEKGFEGRKEGIAVQAVLIE 156  
 DB 354 LDQRILVHEHROPAAVAAEHAVILLE 377

## RESULT 7

; US-09-252-991A-32082  
 ; Sequence 32082, Application US/09252991A  
 ; Patent No. 6551795  
 ; GENERAL INFORMATION:  
 ; APPLICANT: Marc J. Rubenfield et al.  
 ; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO PSEUDOMONAS  
 ; FILE REFERENCE: 107196,136  
 ; CURRENT APPLICATION NUMBER: US/09/252,991A  
 ; PRIOR FILING DATE: 1999-02-18  
 ; PRIOR APPLICATION NUMBER: US 60/074,788  
 ; PRIOR FILING DATE: 1998-02-18  
 ; PRIOR APPLICATION NUMBER: US 60/094,190  
 ; PRIOR FILING DATE: 1998-07-27  
 ; NUMBER OF SEQ ID NOS: 33142  
 ; SEQ ID NO 32082  
 ; LENGTH: 708  
 ; TYPE: PR1  
 ; ORGANISM: Pseudomonas aeruginosa  
 ; US-09-252-991A-32082

Query Match 9.7%; Score 78; DB 4; Length 708;  
 Best Local Similarity 27.8%; Pred. No. 1;  
 Matches 42; Conservative 25; Mismatches 58; Indels 26; Gaps 10;

QY 21 LGGVKI-----PYEKGLE--AHSDDGVVL--HALADALIGAAALGD--IGKHPDTPDK 70  
 DB 206 VGVVGVVAVNGPHATGIDPAHAHGVADVAGPEAGAAVGVGGGGGGLFVLEGGDSDBR 265  
 QY 71 GADSRVLRHYVYGIYK--EKGYKLVNADVTIIAQ--APKMLPHVP--G--MRANIADLETDVD 126  
 DB 266 ABD--FLTEHAHVAWVLEGG-----RLDVVAAGQVAFELLDAAGBQLGAFLLGDVEVGBD 319  
 QY 127 FINVXATTEKLGFEERKEGIAVQAVLIER 157  
 DB 320 LVL-----LGGSLGADHGLGVQAVVAAADR 344

## RESULT 8

; US-09-206-942-71  
 ; Sequence 71, Application US/09206942  
 ; Patent No. 6432659  
 ; GENERAL INFORMATION:  
 ; APPLICANT: Loomore, Sheena M.  
 ; APPLICANT: Yang, Yan-Ping

; APPLICANT: Klein, Michel H.  
 ; TITLE OF INVENTION: Protective Recombinant Haemophilus Influenzae High  
 ; FILE REFERENCE: 1038-861 M15.1b  
 ; CURRENT APPLICATION NUMBER: US/09/206,942  
 ; CURRENT FILING DATE: 1998-12-08  
 ; EARLIER APPLICATION NUMBER: 09/167,568  
 ; EARLIER FILING DATE: 1998-10-07  
 ; NUMBER OF SEQ ID NOS: 95  
 ; SOFTWARE: Patent In Ver. 2.1  
 ; SEQ ID NO 71  
 ; LENGTH: 1477  
 ; TYPE: PR1  
 ; ORGANISM: Haemophilus influenzae  
 ; US-09-206-942-71

Query Match 9.6%; Score 77; DB 4; Length 1477;  
 Best Local Similarity 37.2%; Pred. No. 4.2;  
 Matches 29; Conservative 7; Mismatches 30; Indels 12; Gaps 3;

QY 29 EKGLEASHSDGVDVHLALADALIGAAALGDICK--HFPDTPDKGADSRVLRHYVYGIYK 86  
 DB 385 EKGRAIYWDI---ALIDENINAGSGDIKGTGFTSGHYLSIDN-----ALVK 434  
 QY 87 EKGKLVNADVTIIAQAAP 104  
 DB 435 TKEWLDDPDVTLERBDP 452

## RESULT 9

; US-09-252-991A-18149  
 ; Sequence 18149, Application US/09252991A  
 ; Patent No. 6551795  
 ; GENERAL INFORMATION:  
 ; APPLICANT: Marc J. Rubenfield et al.  
 ; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO PSEUDOMONAS  
 ; FILE REFERENCE: 107196,136  
 ; CURRENT APPLICATION NUMBER: US/09/252,991A  
 ; PRIOR FILING DATE: 1999-02-18  
 ; PRIOR APPLICATION NUMBER: US 60/074,788  
 ; PRIOR FILING DATE: 1998-02-18  
 ; PRIOR APPLICATION NUMBER: US 60/094,190  
 ; PRIOR FILING DATE: 1998-07-27  
 ; NUMBER OF SEQ ID NOS: 33142  
 ; SEQ ID NO 18149  
 ; LENGTH: 220  
 ; TYPE: PR1  
 ; ORGANISM: Pseudomonas aeruginosa  
 ; US-09-252-991A-18149

Query Match 9.4%; Score 75.5; DB 4; Length 220;  
 Best Local Similarity 24.7%; Pred. No. 0.36;  
 Matches 43; Conservative 25; Mismatches 67; Indels 39; Gaps 10;

QY 9 DVHRFNDGDHILGGVXIPYK-----LEASHDG-----DVLHA--LADALIGAAA 54  
 DB 46 DVLGGDDADHL---AIPVDHADTALLLEVHQIGRGRARRHRIHLAAGLAETILGGAA 102  
 QY 55 LGDIGKHPDTPDKGADSRVLRHYVYGIYKKGKYLVDVTIIAQAAPKMLPHVPGR 114  
 DB 103 LGQARDLAHVHDFDVLVAVVDRQA--GVA--GGAQLDDLDLEFFVVEVDAF----- 151  
 QY 115 ANIAADLE--TDVDFINVKATTE-----KLGFEGRKEGI---AVQAVVLIER 157  
 DB 152 DVAARDHGVDRDLGVQQAEGHYMPVDLAGLGVVGRKGLFVAALAAVVLFR 205

## RESULT 10

; US-09-252-991A-20039  
 ; Sequence 20039, Application US/09252991A  
 ; Patent No. 6551795  
 ; GENERAL INFORMATION:

APPLICANT: Marc J. Rubenfield et al.  
TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO PSEUDOMONAS  
FILE REFERENCE: 107196.136  
CURRENT APPLICATION NUMBER: US/09/252,991A  
CURRENT FILING DATE: 1999-02-18  
PRIOR APPLICATION NUMBER: US 60/074,788  
PRIOR FILING DATE: 1998-02-18  
PRIOR APPLICATION NUMBER: US 60/094,190  
PRIOR FILING DATE: 1998-07-27  
NUMBER OF SEQ ID NOS: 33142  
SEQ ID NO 20039  
LENGTH: 400  
TYPE: PRT  
ORGANISM: Pseudomonas aeruginosa  
US-09-252-991A-20039

Query Match 9.4%; Score 75.5; DB 4; Length 400;  
Best Local Similarity 23.8%; Pred. No. 0.89;  
Matches 44; Conservative 27; Mismatches 57; Indels 57; Gaps 9;

QY 2 IRMGVDYHRFPDGHIIIGVYKPKLEAH-----SDGVVLHALLD 47  
DB 121 VDVGVGGEAA--GQHPLIG--QHLLAQHREFRQTPGLDAARLGGAVVEDTLD 173  
QY 48 AIIQAALGDIGKHPDTPNFKADSRVLRHYGVK-----EKGVKLVNADV 97  
DB 174 PLAADLAVGVGQ-----DRVILQRDVLDVETVNHPLADLFRATGALVHRM 221  
QY 98 -----TTIAQAPKMLPHVPCPMANIAADLETVDYI--NKATTTT--KLGFEKX 144  
DB 222 ERNVADVVAIVAQRFEGALHGTAAHV--GPLQDLVHAIPHFDATGIFQGAFGRSQVE 280  
QY 145 EGIAY 149  
DB 281 DGIQV 285

RESULT 11  
US-09-252-991A-23689  
Sequence 23689, Application US/09252991A  
Patent No. 6551795  
GENERAL INFORMATION:  
APPLICANT: Marc J. Rubenfield et al.  
TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO PSEUDOMONAS  
FILE REFERENCE: 107196.136  
CURRENT APPLICATION NUMBER: US/09/252,991A  
CURRENT FILING DATE: 1999-02-18  
PRIOR APPLICATION NUMBER: US 60/074,788  
PRIOR FILING DATE: 1998-02-18  
PRIOR APPLICATION NUMBER: US 60/094,190  
PRIOR FILING DATE: 1998-07-27  
NUMBER OF SEQ ID NOS: 33142  
SEQ ID NO 23689  
LENGTH: 898  
TYPE: PRT  
ORGANISM: Pseudomonas aeruginosa  
US-09-252-991A-23689

Query Match 9.1%; Score 73.5; DB 4; Length 898;  
Best Local Similarity 22.7%; Pred. No. 5.3;  
Matches 44; Conservative 22; Mismatches 67; Indels 61; Gaps 8;

QY 4 VEMGYIV-----HRPVD-----GDHIIIGVYKPKLE-----AHSDGVVLHA 44  
DB 568 VGGDEARVYVMMRRDPRRAVGRGHIAAAGVLLVHQVEVDPVEAHESIAQAGLRP 627  
QY 45 LA-----DALIGAALGDIGKHPDTPD-----PNFKADSRVLRHYVG 83  
DB 628 LAQLAVERRRAADLDSAGHDALVAAAGDALIHHPDTPQAGAGFLGAGPLIV----- 682  
QY 84 IYKKGIVLVNADV-----TTIAQAPKMLPHVPCGA-----NIADLETVDVFINVK 131

DB 683 -----LGHQLADPQAADAMVQLPRGLGEGQRRAVDDDAVEAGGLVVDHETAAHRVTLA 738  
QY 132 ATTEKLGFEKKE 145  
DB 739 AADLQAGSVEGTED 752

RESULT 12  
US-09-252-991A-20237  
Sequence 20237, Application US/09252991A  
Patent No. 6551795  
GENERAL INFORMATION:  
APPLICANT: Marc J. Rubenfield et al.  
TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO PSEUDOMONAS  
FILE REFERENCE: 107196.136  
CURRENT APPLICATION NUMBER: US/09/252,991A  
CURRENT FILING DATE: 1999-02-18  
PRIOR APPLICATION NUMBER: US 60/074,788  
PRIOR FILING DATE: 1998-02-18  
PRIOR APPLICATION NUMBER: US 60/094,190  
PRIOR FILING DATE: 1998-07-27  
NUMBER OF SEQ ID NOS: 33142  
SEQ ID NO 20237  
LENGTH: 1388  
TYPE: PRT  
ORGANISM: Pseudomonas aeruginosa  
US-09-252-991A-20237

Query Match 9.1%; Score 73.5; DB 4; Length 1388;  
Best Local Similarity 24.2%; Pred. No. 10;  
Matches 36; Conservative 20; Mismatches 42; Indels 51; Gaps 7;

QY 17 DHIIIGVYKPKLE-----EAKS-DGCVLH----- 43  
DB 273 DEDAFGRVAVYHGIFFLHYEALQIAGVLEVAADHHAIDGDPVHQRTFGAGALTVDRS 332  
QY 44 --ALADAIIQAALGDIGKHPDTPNFKADSRVLRHYGVYKPKLVNADVTTI 100  
DB 333 QAGLLAAGLASAMLAIEVEGSAV-----AQLRRRVGVV--GIAVAVADVGLA 381  
QY 101 AQAAPKMLPHV-----PCPMANIAADLETVD 126  
DB 382 HQAPVEGLHIAAPG-RAGVERADLAAYE 409

RESULT 13  
US-09-134-001C-3855  
Sequence 3855, Application US/09134001C  
Patent No. 6380370  
GENERAL INFORMATION:  
APPLICANT: Lynn Doucette-Stamm et al  
TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO STAPHYLOCOCCUS  
FILE REFERENCE: GTC-607  
CURRENT APPLICATION NUMBER: US/09/134,001C  
CURRENT FILING DATE: 1998-08-13  
PRIOR APPLICATION NUMBER: US 60/064,964  
PRIOR FILING DATE: 1997-11-08  
PRIOR APPLICATION NUMBER: US 60/055,779  
PRIOR FILING DATE: 1997-08-14  
NUMBER OF SEQ ID NOS: 5674  
SEQ ID NO 3855  
LENGTH: 608  
TYPE: PRT  
ORGANISM: Staphylococcus epidermidis  
US-09-134-001C-3855

Query Match 9.1%; Score 73; DB 4; Length 608;  
Best Local Similarity 21.7%; Pred. No. 3.4;  
Matches 28; Conservative 22; Mismatches 33; Indels 46; Gaps 6;

QY 18 HILIGVKIPYE-----KSLHSDGCVLH-----ALADATGAAAG-----56  
 DB 372 HYMSGELVYDIDMTNKKGLFAGECDPSCHGNRLGANSLSATYCGTAGNAIKYE 431  
 QY 57 DICKEHPDTPD-----NEKQADSRVILRHVYGVKEKGYLVNADVT 98  
 DB 432 NVKSTDTLDDSIPEARVKEKERPDHLIMRGTE-----NAVKLHREIG-EIMTANVT 484  
 QY 99 IIAQAPRML 107  
 DB 485 VVARENEKTL 493

## RESULT 14

US-09-268-347-24  
 ; Sequence 24, Application US/09268347  
 ; Patent No. 6335182  
 ; GENERAL INFORMATION:  
 ; APPLICANT: Loosmore, Sheena M.  
 ; TITLE OF INVENTION: RECOMBINANT HAEMOPHILUS INFLUENZAE ADHESIN PROTEINS  
 ; FILE REFERENCE: 1038-860  
 ; CURRENT APPLICATION NUMBER: US/09/268,347  
 ; CURRENT FILING DATE: 1999-03-16  
 ; NUMBER OF SEQ ID NOS: 54  
 ; SOFTWARE: PatentIn Ver. 2.0  
 ; SEQ ID NO 24  
 ; LENGTH: 1002  
 ; TYPE: PRT  
 ; ORGANISM: Haemophilus influenzae  
 US-09-268-347-24

## Query Match

Best Local Similarity 24.8%; Score 73; DB 4; Length 1002;  
 Matches 33; Conservative 16; Mismatches 46; Indels 38; Gaps 6;

QY 3 RYGMGVDFVRFNDSGDI---IILGVKIPYKGG--LEHSDGCVLHIALADATIGAAAGD 57  
 DB 489 KGVGVDGCGATGDTHTDTLVKSGDKVTLLKAGDNLKVKQSGNFTVLDDELGVKSV-- 546  
 QY 58 IGRHPTDTPNFKGADSRVILRHVYGVKEKGYLVNADVTIIAQAPKMLPHVPGKRAMI 117  
 DB 547 ---EFKDTENGANGAST-----KITKDGITI-----TPANDANG 577  
 QY 118 AADLETVDVPINV 130  
 DB 578 AA--AIDADKIV 588

## RESULT 15

US-09-252-991A-28466  
 ; Sequence 28466, Application US/09252991A  
 ; Patent No. 6531795  
 ; GENERAL INFORMATION:  
 ; APPLICANT: Marc J. Rubenfield et al.  
 ; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO PSEUDOMONAS  
 ; FILE REFERENCE: 107196,136  
 ; CURRENT APPLICATION NUMBER: US/09/252,991A  
 ; CURRENT FILING DATE: 1999-02-18  
 ; PRIOR APPLICATION NUMBER: US 60/074,788  
 ; PRIOR FILING DATE: 1998-02-18  
 ; PRIOR APPLICATION NUMBER: US 60/094,190  
 ; PRIOR FILING DATE: 1998-07-27  
 ; NUMBER OF SEQ ID NOS: 33142  
 ; SEQ ID NO 28466  
 ; LENGTH: 267  
 ; TYPE: PRT  
 ; ORGANISM: Pseudomonas aeruginosa  
 US-09-252-991A-28466

Query Match 9.0%; Score 72.5; DB 4; Length 267;  
 Best Local Similarity 23.2%; Pred. No. 1.1;  
 Matches 35; Conservative 21; Mismatches 44; Indels 51; Gaps 8;

QY 41 VHALADAI-IGAAAL-----GDIG-----KHPPD-----64  
 DB 54 VLRRLAAVAGLACALAAABAEPLKIGVYVGPVGDHGWTTQHEIGRELVENHPEDRVKT 113  
 QY 65 --TDPNFKGADSRVILRHV---YGIYKEKGYLVNADVTIIAQAPKM-LPHVPGMR--A 115  
 DB 114 SEVENVAEGQADARVIRRLAKDGYGLVFTTSFGVYNNPTAKVARQFPKVFPHATGYKKDR 173  
 QY 116 NIADLETVDV-----FINVKATTEKLG 140  
 DB 174 MGTGLSRSYEGRYVGGFLAAKMTESHKIGY 204

Search completed: January 29, 2004, 15:57:06  
 Job time : 5.92742 secs

GenCore version 5.1.6  
Copyright (c) 1993 - 2004 Compugen Ltd.

OM protein - protein search, using sw model

Run on: January 29, 2004, 15:54:19 ; Search time 10.7171 Seconds  
(without alignments)  
3044.503 Million cell updates/sec

Title: US-09-941-947A-14  
Perfect score: 806  
Sequence: 1 MIRVMGVDYHRNDGSHII.....LGPGRKGIAVQAVILIER 157

Scoring table:  
BLOSUM62  
Gapop 10.0 , Gapext 0.5

Searched: 789580 seqs, 207824079 residues  
Total number of hits satisfying chosen parameters: 789580

Minimum DB seq length: 0  
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Database :

- Published Applications AA:\*
- 1: /cgn2\_6/ptodata/1/pubppaa/US07\_PUBCOMB.pep.\*
  - 2: /cgn2\_6/ptodata/1/pubppaa/PC7\_NEW\_PUB.pep.\*
  - 3: /cgn2\_6/ptodata/1/pubppaa/US06\_NEW\_PUB.pep.\*
  - 4: /cgn2\_6/ptodata/1/pubppaa/US06\_PUBCOMB.pep.\*
  - 5: /cgn2\_6/ptodata/1/pubppaa/US07\_NEW\_PUB.pep.\*
  - 6: /cgn2\_6/ptodata/1/pubppaa/PC7US\_PUBCOMB.pep.\*
  - 7: /cgn2\_6/ptodata/1/pubppaa/US08\_NEW\_PUB.pep.\*
  - 8: /cgn2\_6/ptodata/1/pubppaa/US09\_PUBCOMB.pep.\*
  - 9: /cgn2\_6/ptodata/1/pubppaa/US09A\_PUBCOMB.pep.\*
  - 10: /cgn2\_6/ptodata/1/pubppaa/US09B\_PUBCOMB.pep.\*
  - 11: /cgn2\_6/ptodata/1/pubppaa/US09C\_PUBCOMB.pep.\*
  - 12: /cgn2\_6/ptodata/1/pubppaa/US09\_NEW\_PUB.pep.\*
  - 13: /cgn2\_6/ptodata/1/pubppaa/US10A\_PUBCOMB.pep.\*
  - 14: /cgn2\_6/ptodata/1/pubppaa/US10B\_PUBCOMB.pep.\*
  - 15: /cgn2\_6/ptodata/1/pubppaa/US10C\_PUBCOMB.pep.\*
  - 16: /cgn2\_6/ptodata/1/pubppaa/US10D\_NEW\_PUB.pep.\*
  - 17: /cgn2\_6/ptodata/1/pubppaa/US60\_NEW\_PUB.pep.\*
  - 18: /cgn2\_6/ptodata/1/pubppaa/US60\_PUBCOMB.pep.\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	806	100.0	157	10	US-09-934-903-10 Sequence 10, Appl
2	806	100.0	157	10	US-09-934-868-66 Sequence 66, Appl
3	806	100.0	157	11	US-09-941-947A-14 Sequence 14, Appl
4	496	61.5	170	15	US-10-174-410-12 Sequence 12, Appl
5	360.5	44.7	223	12	US-10-259-194A-162 Sequence 162, Appl
6	251	31.1	174	15	US-10-156-761-11502 Sequence 11502, A
7	247.5	30.7	158	12	US-10-128-713A-10 Sequence 10, Appl
8	213.5	26.5	160	10	US-09-738-626-6421 Sequence 6421, Ap
9	153	19.0	32	15	US-10-174-410-135 Sequence 135, Ap
10	153	19.0	199	12	US-10-289-762-585 Sequence 585, App
11	143	17.7	32	15	US-10-174-410-129 Sequence 129, App
12	138	17.1	32	15	US-10-174-410-133 Sequence 133, App
13	138	17.1	32	15	US-10-174-410-134 Sequence 134, App
14	138	17.1	32	15	US-10-174-410-136 Sequence 136, App
15	138	17.1	32	15	US-10-174-410-149 Sequence 149, App

16	133	16.5	32	15	US-10-174-410-125 Sequence 125, App
17	127	15.8	32	15	US-10-174-410-131 Sequence 131, App
18	126	15.6	32	15	US-10-174-410-127 Sequence 127, App
19	126	15.6	32	15	US-10-174-410-150 Sequence 150, App
20	125	15.5	32	15	US-10-174-410-128 Sequence 128, App
21	124	15.4	32	15	US-10-174-410-122 Sequence 122, App
22	124	15.4	32	15	US-10-174-410-140 Sequence 140, App
23	123	15.3	32	15	US-10-174-410-121 Sequence 121, App
24	123	15.3	32	15	US-10-174-410-124 Sequence 124, App
25	119	14.8	32	15	US-10-174-410-148 Sequence 148, App
26	116	14.4	32	15	US-10-174-410-132 Sequence 132, App
27	116	14.3	32	15	US-10-174-410-153 Sequence 153, App
28	115	14.3	32	15	US-10-174-410-130 Sequence 130, App
29	115	14.3	32	15	US-10-174-410-138 Sequence 138, App
30	115	14.3	32	15	US-10-174-410-134 Sequence 134, App
31	114	14.1	32	15	US-10-174-410-139 Sequence 139, App
32	114	14.1	32	15	US-10-174-410-142 Sequence 142, App
33	114	14.1	32	15	US-10-174-410-143 Sequence 143, App
34	114	14.1	32	15	US-10-174-410-156 Sequence 156, App
35	112	13.9	32	15	US-10-174-410-151 Sequence 151, App
36	112	13.9	32	15	US-10-174-410-223 Sequence 223, App
37	110	13.6	32	15	US-10-174-410-123 Sequence 123, App
38	109	13.5	32	15	US-10-174-410-158 Sequence 158, App
39	109	13.5	32	15	US-10-174-410-229 Sequence 229, App
40	109	13.5	32	15	US-10-174-410-230 Sequence 230, App
41	108	13.4	32	15	US-10-174-410-189 Sequence 189, App
42	106	13.2	32	15	US-10-174-410-126 Sequence 126, App
43	106	13.2	32	15	US-10-174-410-146 Sequence 146, App
44	106	13.2	32	15	US-10-174-410-155 Sequence 155, App
45	106	13.2	32	15	US-10-174-410-173 Sequence 173, App

ALIGNMENTS

RESULT 1  
US-09-934-903-10  
Sequence 10, Application US/09934903  
Patent No. US20020102690A1  
GENERAL INFORMATION:  
APPLICANT: Koffas, Mattheos  
APPLICANT: Odum, J. Martin  
APPLICANT: Schenzle, Andreas J.  
APPLICANT: No. US20020102690A1ton, Kelley C.  
APPLICANT: Tomb, Jean-Francois  
APPLICANT: Rouviere, Pierre  
APPLICANT: Piccaraggio, Stephen  
APPLICANT: Cheng, Qiong  
TITLE OR INVENTION: Genes Involved in Isoprenoid Compounds Production  
FILE REFERENCE: C1646 US NA  
CURRENT APPLICATION NUMBER: US/09/934, 903  
CURRENT FILING DATE: 2001-08-22  
PRIOR APPLICATION NUMBER: 60/229, 907  
PRIOR FILING DATE: September 1, 2001  
NUMBER OF SEQ ID NOS: 24  
SOFTWARE: Microsoft Office 97  
SEQ ID NO 10  
LENGTH: 157  
TYPE: PRT  
ORGANISM: Methylobionas 16a  
FEATURE:  
OTHER INFORMATION: Amino acid sequences encoded by ORFs  
US-09-934-903-10

Query Match 100.0%; Score 806; DB 10; Length 157;  
Best Local Similarity 100.0%; Pred. No. 3.5e-81;  
Matches 157; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
QY 1 MIRVMGVDYHRNDGSHIIIGGVKIPYKKGIAHSDDGVVJHALADATIGAAADIGK 60  
DB 1 MIRVMGVDYHRNDGSHIIIGGVKIPYKKGIAHSDDGVVJHALADATIGAAADIGK 60  
QY 61 HPPDTPNFKGADSRVILRHVYGIYKKGKYLWADVTIIAQAQKLVPHVGRANIAAD 120

Db 61 HPPDTPNEFKGADSRVLLRHVYGIYKKGKYLVDVDTIIIAQAPKMLPHVPGKRNATIAAD 120  
Qy 121 LETVDVFINVKATTTTEKLGFEGRKEGIAVOAVVLIER 157  
Db 121 LETVDVFINVKATTTTEKLGFEGRKEGIAVOAVVLIER 157

## RESULT 2

US-09-934-868-66  
Sequence 66, Application US/09934868  
Patent No. US20020137190A1  
GENERAL INFORMATION:  
APPLICANT: Kofinas, Mattheos  
APPLICANT: Odem, James M  
APPLICANT: Schenzle, Andreas J  
TITLE OF INVENTION: IDENTIFYING METHANOTROPHIC BACTERIAL STRAIN  
FILE REFERENCE: CIL596 US NA  
CURRENT APPLICATION NUMBER: US/09/934,868  
CURRENT FILING DATE: 2001-08-22  
PRIOR APPLICATION NUMBER: 60/229,858  
PRIOR FILING DATE: 2000-09-01  
NUMBER OF SEQ ID NOS: 81  
SOFTWARE: Microsoft Office 97  
SEQ ID NO 66  
LENGTH: 157  
TYPE: PRT  
ORGANISM: Methylobionas 16a  
FEATURE:  
OTHER INFORMATION: Amino acid sequences encoded by ISPP  
US-09-934-868-66

Query Match 100.0%; Score 806; DB 10; Length 157;  
Best Local Similarity 100.0%; Pred. No. 3.5e-81;  
Matches 157; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 MIRVGMGYDVHRRNDGDHIIILGCVKIPYEKGLAHSDDGVVLAALDAIIGAALGDIGK 60  
Db 1 MIRVGMGYDVHRRNDGDHIIILGCVKIPYEKGLAHSDDGVVLAALDAIIGAALGDIGK 60  
Qy 61 HPPDTPNEFKGADSRVLLRHVYGIYKKGKYLVDVDTIIIAQAPKMLPHVPGKRNATIAAD 120  
Db 61 HPPDTPNEFKGADSRVLLRHVYGIYKKGKYLVDVDTIIIAQAPKMLPHVPGKRNATIAAD 120  
Qy 121 LETVDVFINVKATTTTEKLGFEGRKEGIAVOAVVLIER 157  
Db 121 LETVDVFINVKATTTTEKLGFEGRKEGIAVOAVVLIER 157

## RESULT 3

US-09-941-947a-14  
Sequence 14, Application US/09941947a  
Publication No. US20030003528A1  
GENERAL INFORMATION:  
APPLICANT: Brzostowicz, Patricia C.  
APPLICANT: Cheng, Qiong  
APPLICANT: DiCosimo, Deana J.  
APPLICANT: Kofinas, Mattheos  
APPLICANT: Miller, Edward S. Jr.  
APPLICANT: Odem, J. Martin  
APPLICANT: Picataggio, Steve  
APPLICANT: Rouviere, Pierre E.  
TITLE OF INVENTION: CAROTENOID PRODUCTION FROM A SINGLE CARBON SOURCE  
FILE REFERENCE: CIL903 US NA  
CURRENT APPLICATION NUMBER: US/09/941,947a  
CURRENT FILING DATE: 2001-09-01  
PRIOR APPLICATION NUMBER: 60/229,907  
PRIOR FILING DATE: 2000-09-01  
PRIOR APPLICATION NUMBER: 60/229,858  
PRIOR FILING DATE: 2000-09-01  
NUMBER OF SEQ ID NOS: 60  
SOFTWARE: Microsoft Office 97  
SEQ ID NO 14

LENGTH: 157  
TYPE: PRT  
ORGANISM: Methylobionas 16a  
US-09-941-947a-14

Query Match 100.0%; Score 806; DB 11; Length 157;  
Best Local Similarity 100.0%; Pred. No. 3.5e-81;  
Matches 157; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 MIRVGMGYDVHRRNDGDHIIILGCVKIPYEKGLAHSDDGVVLAALDAIIGAALGDIGK 60  
Db 1 MIRVGMGYDVHRRNDGDHIIILGCVKIPYEKGLAHSDDGVVLAALDAIIGAALGDIGK 60  
Qy 61 HPPDTPNEFKGADSRVLLRHVYGIYKKGKYLVDVDTIIIAQAPKMLPHVPGKRNATIAAD 120  
Db 61 HPPDTPNEFKGADSRVLLRHVYGIYKKGKYLVDVDTIIIAQAPKMLPHVPGKRNATIAAD 120  
Qy 121 LETVDVFINVKATTTTEKLGFEGRKEGIAVOAVVLIER 157  
Db 121 LETVDVFINVKATTTTEKLGFEGRKEGIAVOAVVLIER 157

## RESULT 4

US-10-174-410-12  
Sequence 12, Application US/10174410  
Publication No. US20030073134A1  
GENERAL INFORMATION:  
APPLICANT: Louie, Gordon V.  
APPLICANT: Buchanan, Sean Grant  
APPLICANT: Gajiwala, Ketan S.  
APPLICANT: Sauder, J. Michael  
TITLE OF INVENTION: CRYSTALS AND STRUCTURES OF  
FILE REFERENCE: 2C-METHYL-D-ERYTHRITOL, 2,4-CYCLODIPHOSPHATE SYNTHASE MECPs  
CURRENT APPLICATION NUMBER: US/10/174,410  
CURRENT FILING DATE: 2002-06-17  
PRIOR APPLICATION NUMBER: 60/259,058  
PRIOR FILING DATE: 2001-06-18  
NUMBER OF SEQ ID NOS: 336  
SOFTWARE: FastSeq for Windows Version 4.0  
SEQ ID NO 12  
LENGTH: 170  
TYPE: PRT  
ORGANISM: Haemophilus influenzae  
FEATURE:  
NAME/KEY: VARIANT  
LOCATION: 1, 69, 108, 115  
OTHER INFORMATION: Xaa = Selenomethionine  
US-10-174-410-12

Query Match 61.5%; Score 496; DB 15; Length 170;  
Best Local Similarity 61.8%; Pred. No. 7.7e-47;  
Matches 97; Conservative 21; Mismatches 39; Indels 0; Gaps 0;

Qy 1 MIRVGMGYDVHRRNDGDHIIILGCVKIPYEKGLAHSDDGVVLAALDAIIGAALGDIGK 60  
Db 3 LIRIGCFVHAHGBRPLIGVVEVPHRTGPIAHSDDGVVLAALDAIIGAALGDIGK 62  
Qy 61 HPPDTPNEFKGADSRVLLRHVYGIYKKGKYLVDVDTIIIAQAPKMLPHVPGKRNATIAAD 120  
Db 63 LPPDTPNEFKGADSRVLLRHVYGIYKKGKYLVDVDTIIIAQAPKMLPHVPGKRNATIAAD 122  
Qy 121 LETVDVFINVKATTTTEKLGFEGRKEGIAVOAVVLIER 157  
Db 123 LQCDIEGVNVLATTTTEKLGFTGDQEGIAECVAVALLIR 159

## RESULT 5

US-10-259-194a-162  
Sequence 162, Application US/10259194a  
Publication No. US20040010815A1  
GENERAL INFORMATION:  
APPLICANT: Lange, Markus B.

```

Query Match      31.1%; Score 251; DB 15; Length 174;
Best Local Similarity 37.6%; Pred. No. 1e-19;
Matches 59; Conservative 23; Mismatches 67; Indels 8; Gaps 3

QY      4 VQMGVDYVRFENDGHIILIGVKKIPEY-KGLEAHSGDDVDVLTALADALIGAAALGIGGF 62
      |||:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|
DB      21 VGIQTDIHAFEDGRHMCWAGLKWEGEGFGLAGHSDADVAAALCNALPSAAGLGDIGGF 80

QY      63 PDDTPNFKGADSRVLAEHVIGIVVEKGYKLVNADVTLIAQAPKYLPHVPGKRNIAADLE 122
      |||:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|
DB      81 GTRPPESGASGVTLTLEARNIVAAAGTIGNVAVQVGGPPKLI---GRRDEAQKIL 135

QY      123 TDV--DEINNVKATTEKLGFEGRREGIAVQAVLIER 157
      ||:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|
DB      136 SDVAGAPVSVSGATTTDGLGFGFRDEGLMAVATLVVR 172

RESULT 7
US-10-128-713A-10
; Sequence 10, Application US/10128713A
; Publication No. US20030170847A1
; GENERAL INFORMATION:
; APPLICANT: Bramucci, Michael G
; TITLE OR INVENTION: Genes Involved in Isoprenoid Compound Production
; EMBL REFERENCE: CC-1788
; CURRENT APPLICATION NUMBER: US/10/128, 713A
; CURRENT FILING DATE: 2002-04-22
; NUMBER OF SEQ ID NOS: 36

```

```

: SEQ ID NO: 10
: LENGTH: 158
: TYPE: PRT
: ORGANISM: Rhodococcus erythropolis
: US-10-128-713A-10

Query Match      30.7%; Score 247.5; DB: 12; Length 158;
Best Local Similarity 37.8%; Pred. No. 2.2e-19;
Matches 59; Conservative 24; Mismatches 66; Indels 7; Gaps 2.

Oy      2  IRVMSGYDVARFMDGDHIIILGYSKIPYEKGLFAHSDDGVTLHALADAILGAALGDIQKH 61
      |||:|||||:|:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db      1  MRVGLGTDVPIEVGRRCWMAAGLLFELADDCSGHSDDVAVHALCDLLSAGIGDLSGV 60

Oy      62  PPDTPNFKGADSRVILLRHVYGLVKEKGYLVANADVTIIAQPRLPHVPMGRANIAADL 121

```

0Y 122 ETDV--DFINVKATTEKLGFBGRKEGIAVOAVYLI 155  
 Db 116 LSLDLAGAPVSASATTDGLSLTGRGBEGLAMATV 151  
 RESULT 8  
 US-09-738-626-6421  
 ; Sequence 6421, Application US/09738626  
 ; Publication No. US20020197605A1  
 ; GENERAL INFORMATION:  
 ; APPLICANT: NAKAGAWA, SATOSHI  
 ; APPLICANT: MIZOGUCHI, HIROSHI  
 ; APPLICANT: ANDO, SEIKO  
 ; APPLICANT: HAYASHI, MIKIRO  
 ; APPLICANT: OCHIAI, KEIKO  
 ; APPLICANT: YOKOI, HARUHIKO  
 ; APPLICANT: TATEISHI, NAOKO  
 ; APPLICANT: SENOH, AKIHITO  
 ; APPLICANT: IKEDA, MASATO  
 ; APPLICANT: OZAKI, AKIO  
 ; TITLE OF INVENTION: NOVEL POLYNUCLEOTIDES  
 ; FILE REFERENCE: 249-125  
 ; CURRENT APPLICATION NUMBER: US/09/738, 626  
 ; CURRENT FILING DATE: 2000-12-18  
 ; PRIOR APPLICATION NUMBER: JP 99/377484

```

; PRIOR APPLICATION NUMBER: JP 00/159162
; PRIOR FILING DATE: 2000-04-07
; PRIOR APPLICATION NUMBER: JP 00/280988
; PRIOR FILING DATE: 2000-08-03
; NUMBER OF SEQ ID NOS: 7059
; SOFTWARE: PatentIn ver. 3.0
; SEQ ID NO 6421
; LENGTH: 160
; TYPE: PRT
; ORGANISM: Corynebacterium glutamicum
US-09-738-626-6421

```

```

Query Match      26.5%; Score 213.5; DB 10; Length 160;
Best Local Similarity 32.9%; Pred. No. 1,3e-15;
Matches 51; Conservative 25; Mismatches 72; Indels 7; Gaps 2;

```

```

QY 3 RVGAGYVHRFNDGHIILGVKIPYKGLSHSDGVVLAALADALIGAAALGIGKHF 62
DB 8 RVGIATDAHQIBAGKPCWIACLLFEGVDCGEGSDGVVAALIVDALISASGLGGSFV 67
QY 63 PPTDPFKAGDSRVLLRHYYGIYKGYKLVNADVTIIIAQPKMLPHVPGKANTADLE 122
DB 66 GVGREPYDVSGTQILKEVRELSAHGYIGVMAAQLVGQTKF-----GPRRERAGQVT 122
QY 123 TDV--DFINVKATTEKLGFBGRKKGIAVQAVLI 155
DB 123 SEIIGAPCISLAFTTDMGFTGRSGSRASVATNAV 157

```

```

RESULT 9
US-10-174-410-135
; Sequence 135, Application US/10174410
; Publication No. US20030073134A1
; GENERAL INFORMATION:
; APPLICANT: Louie, Gordon V.
; APPLICANT: Buchanan, Sean Grant
; APPLICANT: Gajiwala, Ketan S.
; APPLICANT: Sauder, J. Michael
; TITLE OF INVENTION: CRYSTALS AND STRUCTURES OF
; FILE REFERENCE: 52498200300
; CURRENT APPLICATION NUMBER: US/10/174,410
; CURRENT FILING DATE: 2002-06-17
; PRIOR APPLICATION NUMBER: 60/299,058
; PRIOR FILING DATE: 2001-06-18
; NUMBER OF SEQ ID NOS: 336
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 135
; LENGTH: 32
; TYPE: PRT
; ORGANISM: Pseudomonas aeruginosa
US-10-174-410-135

```

```

Query Match      19.0%; Score 153; DB 15; Length 32;
Best Local Similarity 87.5%; Pred. No. 7e-10;
Matches 28; Conservative 3; Mismatches 1; Indels 0; Gaps 0;

```

```

QY 40 VVLAALADALIGAAALGIGKHPPTDPNFKG 71
DB 1 VVLAALSDALIGAAALGIGKHPPTDPNFKG 32

```

```

RESULT 10
US-10-289-762-585
; Sequence 585, Application US/10289762
; Publication No. US2004006218A1
; GENERAL INFORMATION:
; APPLICANT: Griffiths, R.
; TITLE OF INVENTION: Chlamydia pneumoniae genomic sequence and polypeptides, fragments
; FILE REFERENCE: thereof and uses thereof, in particular for the diagnosis, prevention
; CURRENT APPLICATION NUMBER: US/10/289,762

```

```

; CURRENT FILING DATE: 2003-03-27
; NUMBER OF SEQ ID NOS: 6849
; SEQ ID NO 585
; LENGTH: 199
; TYPE: PRT
; ORGANISM: Chlamydia pneumoniae
US-10-289-762-585

```

```

Query Match      19.0%; Score 153; DB 12; Length 199;
Best Local Similarity 29.4%; Pred. No. 8.5e-09;
Matches 47; Conservative 26; Mismatches 75; Indels 12; Gaps 5;

```

```

QY 3 RVGAGYVHRF---NDGDHIIILGVKIPYKGLSHSDGVVLAALADALIGAA---ALG 56
DB 40 RTGIGDSHRFLPSSSTKPCILGSIYDPCGQANSDDGIIFRAICNAISVTNKILG 99
QY 57 DIGRHPDTPNFKG-ADSRVLLRHYYGIYKGYKLVNADVTIIIAQPKMLPHVPGKRA 115
DB 100 KVADELQOT---RGITDSGIYIEALKSIK-PQKISHVAITFGSRPRFLCTSLARQ 154
QY 116 NIADLETDVDFINVKATTEKLGFBGRKKGIAVQAVLI 155
DB 155 NIAQVMULTPTDITGITATSGGLSDFCGSGVQCFCVLTV 194

```

```

RESULT 11
US-10-174-410-129
; Sequence 129, Application US/10174410
; Publication No. US20030073134A1
; GENERAL INFORMATION:
; APPLICANT: Louie, Gordon V.
; APPLICANT: Buchanan, Sean Grant
; APPLICANT: Gajiwala, Ketan S.
; APPLICANT: Sauder, J. Michael
; TITLE OF INVENTION: CRYSTALS AND STRUCTURES OF
; FILE REFERENCE: 52498200300
; CURRENT APPLICATION NUMBER: US/10/174,410
; CURRENT FILING DATE: 2002-06-17
; PRIOR APPLICATION NUMBER: 60/299,058
; PRIOR FILING DATE: 2001-06-18
; NUMBER OF SEQ ID NOS: 336
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 129
; LENGTH: 32
; TYPE: PRT
; ORGANISM: C. perfringens
US-10-174-410-129

```

```

Query Match      17.7%; Score 143; DB 15; Length 32;
Best Local Similarity 78.1%; Pred. No. 8.9e-09;
Matches 25; Conservative 5; Mismatches 2; Indels 0; Gaps 0;

```

```

QY 40 VVLAALADALIGAAALGIGKHPPTDPNFKG 71
DB 1 VVLAALSDALIGAAALGIGKHPPTDPNFKG 32

```

```

RESULT 12
US-10-174-410-133
; Sequence 133, Application US/10174410
; Publication No. US20030073134A1
; GENERAL INFORMATION:
; APPLICANT: Louie, Gordon V.
; APPLICANT: Buchanan, Sean Grant
; APPLICANT: Gajiwala, Ketan S.
; APPLICANT: Sauder, J. Michael
; TITLE OF INVENTION: CRYSTALS AND STRUCTURES OF
; FILE REFERENCE: 52498200300
; CURRENT APPLICATION NUMBER: US/10/174,410
; CURRENT FILING DATE: 2002-06-17
; PRIOR APPLICATION NUMBER: 60/299,058

```

```

; PRIOR FILING DATE: 2001-06-18
; NUMBER OF SEQ ID NOS: 336
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 133
; LENGTH: 32
; TYPE: PRT
; ORGANISM: S. typhimurium
US-10-174-410-133

Query Match      17.1%; Score 138; DB 15; Length 32;
Best Local Similarity 84.4%; Pred. No. 3.2e-08;
Matches 27; Conservative 1; Mismatches 4; Indels 0; Gaps 0;

Db
40 VVHALLADALGAAALGDIGKHPPTDPAFKG 71
1 VVHALLTDALGAAALGDIGKLPPTDPAFKG 32

RESULT 13
US-10-174-410-134
; Sequence 134, Application US/10174410
; Publication No. US20030073134A1
; GENERAL INFORMATION:
; APPLICANT: Louie, Gordon V.
; APPLICANT: Buchanan, Sean Grant
; APPLICANT: Gajiwala, Ketan S.
; APPLICANT: Sauder, J. Michael
; TITLE OF INVENTION: CRYSTALS AND STRUCTURES OF
; TITLE OF INVENTION: 2C-METHYL-D-ERYTHRITOL, 2,4-CYCLODIPHOSPHATE SYNTHASE MECPs
; FILE REFERENCE: 524982000300
; CURRENT APPLICATION NUMBER: US/10/174,410
; CURRENT FILING DATE: 2002-06-17
; PRIOR APPLICATION NUMBER: 60/299,058
; PRIOR FILING DATE: 2001-06-18
; NUMBER OF SEQ ID NOS: 336
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 134
; LENGTH: 32
; TYPE: PRT
; ORGANISM: S. enterica
US-10-174-410-134

Query Match      17.1%; Score 138; DB 15; Length 32;
Best Local Similarity 84.4%; Pred. No. 3.2e-08;
Matches 27; Conservative 1; Mismatches 4; Indels 0; Gaps 0;

Db
40 VVHALLADALGAAALGDIGKHPPTDPAFKG 71
1 VVHALLTDALGAAALGDIGKLPPTDPAFKG 32

RESULT 14
US-10-174-410-136
; Sequence 136, Application US/10174410
; Publication No. US20030073134A1
; GENERAL INFORMATION:
; APPLICANT: Louie, Gordon V.
; APPLICANT: Buchanan, Sean Grant
; APPLICANT: Gajiwala, Ketan S.
; APPLICANT: Sauder, J. Michael
; TITLE OF INVENTION: CRYSTALS AND STRUCTURES OF
; TITLE OF INVENTION: 2C-METHYL-D-ERYTHRITOL, 2,4-CYCLODIPHOSPHATE SYNTHASE MECPs
; FILE REFERENCE: 524982000300
; CURRENT APPLICATION NUMBER: US/10/174,410
; CURRENT FILING DATE: 2002-06-17
; PRIOR APPLICATION NUMBER: 60/299,058
; PRIOR FILING DATE: 2001-06-18
; NUMBER OF SEQ ID NOS: 336
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 136
; LENGTH: 32
; TYPE: PRT
; ORGANISM: Escherichia coli
```

```

US-10-174-410-136

Query Match      17.1%; Score 138; DB 15; Length 32;
Best Local Similarity 84.4%; Pred. No. 3.2e-08;
Matches 27; Conservative 1; Mismatches 4; Indels 0; Gaps 0;

Db
40 VVHALLADALGAAALGDIGKHPPTDPAFKG 71
1 VVHALLTDALGAAALGDIGKLPPTDPAFKG 32

RESULT 15
US-10-174-410-149
; Sequence 149, Application US/10174410
; Publication No. US20030073134A1
; GENERAL INFORMATION:
; APPLICANT: Louie, Gordon V.
; APPLICANT: Buchanan, Sean Grant
; APPLICANT: Gajiwala, Ketan S.
; APPLICANT: Sauder, J. Michael
; TITLE OF INVENTION: CRYSTALS AND STRUCTURES OF
; TITLE OF INVENTION: 2C-METHYL-D-ERYTHRITOL, 2,4-CYCLODIPHOSPHATE SYNTHASE MECPs
; FILE REFERENCE: 524982000300
; CURRENT APPLICATION NUMBER: US/10/174,410
; CURRENT FILING DATE: 2002-06-17
; PRIOR APPLICATION NUMBER: 60/299,058
; PRIOR FILING DATE: 2001-06-18
; NUMBER OF SEQ ID NOS: 336
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 149
; LENGTH: 32
; TYPE: PRT
; ORGANISM: Escherichia coli
US-10-174-410-149

Query Match      17.1%; Score 138; DB 15; Length 32;
Best Local Similarity 84.4%; Pred. No. 3.2e-08;
Matches 27; Conservative 1; Mismatches 4; Indels 0; Gaps 0;

Db
40 VVHALLADALGAAALGDIGKHPPTDPAFKG 71
1 VVHALLTDALGAAALGDIGKLPPTDPAFKG 32

Search completed: January 29, 2004, 16:21:17
Job time : 10.7171 secs
```

GenCore version 5.1.6  
Copyright (c) 1993 - 2004 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: January 29, 2004, 15:44:49 ; Search time 4.61946 Seconds  
(without alignments)  
3268.453 Million cell updates/sec

Title: US-09-941-947a-14  
Perfect score: 806  
Sequence: 1 MIRVGQYDVHRFNDGDHIT.....LGFGRNKGIAVQAVVLI 157

Scoring table: BLOSUM62  
Gapop 10.0, Gapext 0.5

Searched: 283308 seqs, 9616882 residues

Total number of hits satisfying chosen parameters: 283308

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Database:

1: PIR\_76:.\*  
2: p1r1:.\*  
3: p1r2:.\*  
4: p1r3:.\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

#### SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	575	71.3	157	A83194	conserved hypotet
2	554	68.7	158	D82311	conserved hypotet
3	542	67.2	162	AB0408	2C-methyl-D-erythr
4	537	66.6	159	I55083	2C-methyl-D-erythr
5	537	66.6	159	A85924	2C-methyl-D-erythr
6	537	66.6	159	AD0856	2C-methyl-D-erythr
7	537	66.6	159	H91078	2C-methyl-D-erythr
8	534	66.3	158	F64156	hypothetical prote
9	520	64.5	157	F69741	gltx 5'-region con
10	520	64.5	157	D83663	hypothetical prote
11	505	62.7	160	G81073	YgbB/YacN family p
12	503	62.4	160	S70845	hypothetical prote
13	503	62.4	160	C81867	conserved hypotet
14	452	56.1	176	A82701	conserved hypotet
15	451	56.0	157	AE1104	B. subtilis YacN p
16	443	55.0	157	AE1466	hypothetical prote
17	425	52.7	379	S34980	hypothetical prote
18	424	52.6	155	C96953	hypothetical prote
19	414	51.4	380	A13359	2-C-methyl-D-eryth
20	408	50.6	382	F87464	conserved hypotet
21	403	50.0	161	E84978	hypothetical prote
22	403	50.0	165	AD2291	2-C-methyl-D-eryth
23	403	50.0	371	E81256	hypothetical prote
24	396	49.1	161	S73363	hypothetical prote
25	373	46.3	211	H96664	unknown protein, 3
26	372	46.2	400	AC2754	ispD/ispF bifuncti
27	372	46.2	400	AC2754	ispD/ispF bifuncti
28	345	42.8	409	G71936	hypothetical prote
29	343	42.6	406	D64647	conserved hypotet

30	329	40.8	399	2	G71314	conserved hypotet
31	306.5	38.0	165	2	F72351	conserved hypotet
32	301.5	37.4	156	2	F70382	conserved hypotet
33	289.5	35.9	161	2	F75542	conserved hypotet
34	285.5	35.4	158	2	B86949	hypothetical prote
35	276.5	34.3	159	2	C70607	hypothetical prote
36	269	33.4	240	2	B71615	YgbB protein, 1st
37	210	26.1	178	2	E81672	conserved hypotet
38	202	25.1	178	2	C71514	hypothetical prote
39	153	19.0	176	2	D72065	conserved hypotet
40	153	19.0	176	2	G86558	hypothetical prote
41	104.5	13.0	169	2	S73495	hypothetical prote
42	102.5	12.7	169	2	G64250	plasmid surface ex
43	83	10.3	593	2	T21510	hypothetical prote
44	81.5	10.1	294	2	C83497	ribosome ABC transp
45	81	10.0	434	2	T47545	monodehydroascorba

#### ALIGNMENTS

##### RESULT 1

A83194

conserved hypothetical protein PA3627 [imported] - Pseudomonas aeruginosa (strain PA01)

C/Species: Pseudomonas aeruginosa

C/Date: 15-Sep-2000 #sequence\_revision 15-Sep-2000 #text\_change 09-Dec-2002

C/Accession: A83194

R/Storer, C.K.; Pham, X.Q.; Erwin, A.L.; Mizoguchi, S.D.; Warren, P.; Hickey, M.J.; Br

adman, S.; Yuan, Y.; Brody, L.L.; Coulter, S.N.; Folger, K.R.; Kae, A.; Laroig, K.; Lim,

., Loty, S.; Olson, M.V.

Nature 406, 959-964, 2000

A/Title: Complete genome sequence of Pseudomonas aeruginosa PA01, an opportunistic pathog

A/Reference number: A82950; UID:20437337; PMID:10984043

A/Accession: A83194

A/Status: preliminary

A/Molecule type: DNA

A/Residues: 1-157 <STO>

A/Cross-references: GB:AE004782; GB:AE004091; MID:g9949772; PIDN:AA07015.1; GSPDB:GN001;

A/Experimental source: strain PA01

C/Genetics:

A/Genes: PA3627

C/Superfamily: 2C-methyl-D-erythritol 2,4-cyclodiphosphate synthase

Query Match

Best Local Similarity 71.3%; Score 575; DB 2; Length 157;

Matches 110; Conservative 17; Mismatches 27; Indels 0; Gaps 0;

QY 2 IVVGQYDVHRFNDGDHIIIGSVKIPYKGLKSHSDGVVLAADALGAAGDIGK 61

DB 1 MGVGQYDVHRFNDGDHIIIGSVKIPYKGLKSHSDGVVLAADALGAAGDIGK 60

QY 62 PDGTPNFKGADSRVLLRHVYGVKYGKYLVDVITIAQPKXLPHVPGKANIADL 121

DB 61 PDGTPNFKGADSRVLLRHVYGVKYGKYLVDVITIAQPKXLPHVPGKANIADL 120

QY 122 ETDVDPIVWKAATTEKELGEGKXGIAVQAVVLI 155

DB 121 GVAVDVWKAATTEKELGEGKXGIAVQAVVLI 154

##### RESULT 2

conserved hypothetical protein VC0529 [imported] - Vibrio cholerae (strain N16961 serogr

C/Species: Vibrio cholerae

C/Date: 18-Aug-2000 #sequence\_revision 20-Aug-2000 #text\_change 09-Dec-2002

C/Accession: D82311

R/Heldberg, J.F.; Risen, J.A.; Nelson, W.C.; Clayton, R.A.; Gwin, M.L.; Dodson, R.J.;

chardson, D.; Ermolaeva, M.D.; Vamathevan, J.; Bass, S.; Qin, H.; Dregoli, I.; Sellers, P.

l, R.R.; Mekalanos, J.J.; Venter, J.C.; Fraser, C.M.

Nature 406, 477-483, 2000

A/Title: DNA Sequence of both chromosomes of the cholera pathogen Vibrio cholerae.

A/Reference number: A82035; UID:20406833; PMID:10952301

A/Accession: D82311

A:Status: preliminary  
 A:Molecule type: DNA  
 A:Residues: 1-158 <HEI>  
 A:Cross-references: GB:AE004139; GB:AE003852; MID:9664953; PIDN:AAF93697.1; GSPDB:GN001  
 A:Experimental source: serogroup O1; strain N16961; biotype E1 Tor  
 A:Genetics:  
 A:Gene: VC0529  
 A:Map position: 1  
 C:Superfamily: 2C-methyl-D-erythritol 2,4-cyclodiphosphate synthase

Query Match 68.7%; Score 554; DB 2; Length 158;  
 Best Local Similarity 70.7%; Pred. No. 4.5e-44;  
 Matches 111; Conservative 16; Mismatches 30; Indels 0; Gaps 0;

QY 1 MIRVGKGVYHRRPNDGDHIIIGGVKIPYKGLAEHSDGDVVLHALLADATILGAALGDIGK 60  
 DB 1 MIRIGHGFVHAFGEGPFIITIGGVRIPEYKGLAHSDDGVVLAHALLADATILGAALGDIGK 60  
 QY 61 HPPDTPNFKGADSRVLLRHVYGVKEKGYKLVNADVTIIAQAQKPLPHVPGKRAMIAAD 120  
 DB 61 HPPDTPNFKGADSRVLLRHVYGVKEKGYKLVNADVTIIAQAQKPLPHVPGKRAMIAAD 120  
 QY 121 LEVDVDFINVKATTEKLGFEGRKEGIAVOAVVLI 157  
 DB 121 LEVDVDFINVKATTEKLGFEGRKEGIAVOAVVLI 157

## RESULT 3

AB0408  
 2C-methyl-D-erythritol 2,4-cyclodiphosphate synthase [imported] - Yersinia pestis (strain  
 C:Species: Yersinia pestis  
 C:Date: 02-Nov-2001 #sequence\_revision 02-Nov-2001 #text\_change 09-Dec-2002  
 A:Accession: AB0408  
 R:Parkhill, J.; Wren, B.W.; Thomson, N.R.; Titball, R.W.; Holden, M.T.G.; Prentice, M.B.;  
 deno-Tarraga, A.M.; Chillingworth, T.; Cronin, A.; Davies, R.M.; Davis, P.; Dougan, G.;  
 Nature 413, 523-527, 2001  
 A:Title: Genome sequence of Yersinia pestis, the causative agent of plague.  
 A:Reference number: AB0001; MUID:21470413; PMID:11586360  
 A:Accession: AB0408  
 A:Status: preliminary  
 A:Molecule type: DNA  
 A:Residues: 1-162 <KUR>  
 A:Cross-references: GB:AL590842; PIDN:CAC92590.1; PID:G15981287; GSPDB:GN00175  
 C:Genetics:  
 A:Gene: 189P  
 C:Superfamily: 2C-methyl-D-erythritol 2,4-cyclodiphosphate synthase

Query Match 67.2%; Score 542; DB 2; Length 162;  
 Best Local Similarity 68.6%; Pred. No. 5.9e-43;  
 Matches 107; Conservative 17; Mismatches 30; Indels 2; Gaps 1;

QY 2 IRVGMGVYHRRPNDGDHIIIGGVKIPYKGLAEHSDGDVVLHALLADATILGAALGDIG 59  
 DB 1 MIRIGHGFVHAFGEGPFIITIGGVRIPEYKGLAHSDDGVVLAHALLADATILGAALGDIG 60  
 QY 60 KHPDTPNFKGADSRVLLRHVYGVKEKGYKLVNADVTIIAQAQKPLPHVPGKRAMIAA 119  
 DB 61 KHPDTPNFKGADSRVLLRHVYGVKEKGYKLVNADVTIIAQAQKPLPHVPGKRAMIAA 120  
 QY 120 LEVDVDFINVKATTEKLGFEGRKEGIAVOAVVLI 155  
 DB 121 LEVDVDFINVKATTEKLGFEGRKEGIAVOAVVLI 156

## RESULT 4

155083  
 2C-methyl-D-erythritol 2,4-cyclodiphosphate synthase [similarity] - Escherichia coli (st  
 C:Species: Escherichia coli  
 C:Date: 07-Jun-1996 #sequence\_revision 07-Jun-1996 #text\_change 09-Dec-2002  
 C:Accession: 155083; F55055  
 R:Hi, C.; Ichikawa, J.K.; Ravetto, J.J.; Kuo, H.C.; Fu, J.C.; Clarke, S.  
 J. Bacteriol. 176, 6015-6022, 1994

A:Title: A new gene involved in stationary-phase survival located at 59 minutes on the E  
 A:Reference number: 155083; MUID:95014035; PMID:7928962  
 A:Accession: 155083  
 A:Status: preliminary; translated from GB/EMBL/DBJ  
 A:Molecule type: DNA  
 A:Residues: 1-159 <RES>  
 A:Cross-references: GB:L07942; MID:91036737; PIDN:AAA79837.1; PID:G433711  
 R:Blattner, F.R.; Plunkett III, G.; Bloch, C.A.; Perna, N.T.; Burland, V.; Riley, M.; C  
 A.; Rose, D.J.; Mau, B.; Shao, Y.  
 Science 277, 1453-1462, 1997  
 A:Title: The complete genome sequence of Escherichia coli K-12.  
 A:Reference number: A64720; MUID:97426617; PMID:9278503  
 A:Accession: F55055  
 A:Status: preliminary; nucleic acid sequence not shown; translation not shown

A:Molecule type: DNA  
 A:Residues: 1-159 <BLAT>  
 A:Cross-references: GB:AE00358; GB:U00096; MID:92367156; PIDN:AACT5788.1; PID:G1789103;  
 A:Experimental source: strain K-12, substrain MG1655  
 C:Genetics:  
 A:Gene: y98b  
 C:Superfamily: 2C-methyl-D-erythritol 2,4-cyclodiphosphate synthase

Query Match 66.6%; Score 537; DB 2; Length 159;  
 Best Local Similarity 69.5%; Pred. No. 1.7e-42;  
 Matches 107; Conservative 14; Mismatches 33; Indels 0; Gaps 0;

## RESULT 5

AB5924  
 2C-methyl-D-erythritol 2,4-cyclodiphosphate synthase [similarity] - Escherichia coli (st  
 C:Species: Escherichia coli  
 C:Date: 16-Feb-2001 #sequence\_revision 16-Feb-2001 #text\_change 09-Dec-2002  
 A:Accession: AB5924  
 R:Perna, N.T.; Plunkett III, G.; Burland, V.; Mau, B.; Glasner, J.D.; Rose, D.J.; Mayhew  
 Miller, L.; Grobeck, E.J.; Davis, N.W.; Lam, A.; Dimalanta, E.; Potamousis, K.; Apodaca,  
 Nature 409, 529-533, 2001  
 A:Title: Genome sequence of enterohemorrhagic Escherichia coli O157:H7.  
 A:Reference number: AB5480; MUID:21074935; PMID:11206551  
 A:Accession: AB5924  
 A:Status: preliminary  
 A:Molecule type: DNA  
 A:Residues: 1-159 <STO>  
 A:Cross-references: GB:AE005174; MID:912517200; PIDN:AA657853.1; GSPDB:GN00145; UMGF:24C  
 A:Experimental source: strain O157:H7, substrain EDL933  
 C:Genetics:  
 A:Gene: y98b  
 C:Superfamily: 2C-methyl-D-erythritol 2,4-cyclodiphosphate synthase

Query Match 66.6%; Score 537; DB 2; Length 159;  
 Best Local Similarity 69.5%; Pred. No. 1.7e-42;  
 Matches 107; Conservative 14; Mismatches 33; Indels 0; Gaps 0;

QY 2 IRVGMGVYHRRPNDGDHIIIGGVKIPYKGLAEHSDGDVVLHALLADATILGAALGDIGK 61  
 DB 1 MIRIGHGFVHAFGEGPFIITIGGVRIPEYKGLAHSDDGVVLAHALLADATILGAALGDIGK 60  
 QY 62 FPPDTPNFKGADSRVLLRHVYGVKEKGYKLVNADVTIIAQAQKPLPHVPGKRAMIAADL 121  
 DB 61 FPPDTPNFKGADSRVLLRHVYGVKEKGYKLVNADVTIIAQAQKPLPHVPGKRAMIAADL 120  
 QY 122 EVDVDFINVKATTEKLGFEGRKEGIAVOAVVLI 155



A/Cross-references: GB:299104; GB:AL0009126; NID:G2632267; PIDN:CA81867.1; PID:G2632358  
 A/Experimental source: strain 168  
 R/Ogasawara, N.; Nakai, S.; Yoshikawa, H.  
 DNA Res. 1, 1-14, 1994  
 A/Title: Systematic sequencing of the 180 kbpase region of the *Bacillus subtilis* chrom  
 A/Reference number: S65367; MUID:96051385; PMID:7584024  
 A/Accession: S66120  
 A/Status: nucleic acid sequence not shown; translation not shown  
 A/Molecule type: DNA  
 A/Residues: 1-158 <GAG>  
 A/Cross-references: EMBL:D26185; NID:G467326; PIDN:BA05325.1; PID:G467479  
 A/Note: the nucleotide sequence was submitted to the EMBL Data Library, December 1993  
 R/Gagnon, Y.; Breton, R.; Futerer, H.; Pelchat, M.; Grunberg-Manago, M.; Lapointe, J.  
 J. Biol. Chem. 269, 7473-7482, 1994  
 A/Title: Clustering and co-transcription of the *Bacillus subtilis* genes encoding the am  
 A/Reference number: A53402; MUID:9411772; PMID:7510287  
 A/Accession: A53402  
 A/Molecule type: DNA  
 A/Residues: 137-158 <GAG>  
 A/Cross-references: GB:114580  
 C/Genetics:  
 A/Experimental source: strain C-125  
 A/Status: YACN  
 A/Reference number: 2C-methyl-D-erythritol 2,4-cyclodiphosphate synthase  
 Query Match 66.3%; Score 534; DB 2; Length 158;  
 Best Local Similarity 65.6%; Pred. No. 3.2e-42;  
 Matches 103; Conservative 18; Mismatches 36; Indels 0; Gaps 0;

QY 1 MIRVGKGVYHFNDDHIIIGGVKIPYKGLKSHSDGVVLAHALADALIGAAALGDIK 60  
 DB 1 MFIIGGFPDVHGVKGRPLIGSIRIPYKGLKSHSDAVLHTVADALGAGVBEDIK 60

QY 61 HEPDTPNFKGADSRVLLRHVYGIYKKGKLVNADVTIIAQPXKLPHVPGKRAIIAD 120  
 DB 61 HEPDTPNFKGADSRVLLRHVYGIYKKGKLVNADVTIIAQPXKLPHVPGKRAIIAD 120

QY 121 LETVDVFINVKATTEKLGFEGRKEGIAVQAVLIER 157  
 DB 121 LEADVQVNVKATTEKLGFTGSAAGIAQAVLIER 157

RESULT 10  
 D83663  
 hypotetical protein B910108 (imported) - *Bacillus halodurans* (strain C-125)  
 C/Species: *Bacillus halodurans*  
 C/Date: 01-Dec-2000 #sequence\_revision 01-Dec-2000 #text\_change 09-Dec-2002  
 C/Accession: D83663  
 R/Takami, H.; Nakasone, K.; Takaki, Y.; Maeno, G.; Sasaki, R.; Masui, N.; Fujii, F.; Hira  
 Nucleic Acids Res. 28, 4317-4331, 2000  
 A/Title: Complete genome sequence of the alkaliphilic bacterium *Bacillus halodurans* and  
 A/Reference number: A83650; MUID:20512582; PMID:11058132  
 A/Accession: D83663  
 A/Status: preliminary  
 A/Molecule type: DNA  
 A/Residues: 1-157 <STO>  
 A/Cross-references: GB:AP001507; GB:BA000004; NID:G10172612; PIDN:BA03827.1; GSPDB:GN00  
 A/Experimental source: strain C-125  
 C/Genetics:  
 A/Status: YACN  
 A/Reference number: 2C-methyl-D-erythritol 2,4-cyclodiphosphate synthase  
 Query Match 64.5%; Score 520; DB 2; Length 157;  
 Best Local Similarity 66.5%; Pred. No. 6.2e-41;  
 Matches 103; Conservative 16; Mismatches 36; Indels 0; Gaps 0;

QY 1 MIRVGKGVYHFNDDHIIIGGVKIPYKGLKSHSDGVVLAHALADALIGAAALGDIK 60  
 DB 1 MIRVGKGVYHFNDDHIIIGGVKIPYKGLKSHSDGVVLAHALADALIGAAALGDIK 60

QY 61 HEPDTPNFKGADSRVLLRHVYGIYKKGKLVNADVTIIAQPXKLPHVPGKRAIIAD 120  
 DB 61 HEPDTPNFKGADSRVLLRHVYGIYKKGKLVNADVTIIAQPXKLPHVPGKRAIIAD 120

QY 121 LETVDVFINVKATTEKLGFEGRKEGIAVQAVLIER 155  
 DB 121 LEADVQVNVKATTEKLGFTGSAAGIAQAVLIER 155

RESULT 11  
 G81073  
 ygdB/Yacn family protein NMB1512 (imported) - *Neisseria meningitidis* (strain MCS8 serogr  
 C/Species: *Neisseria meningitidis*  
 C/Date: 31-Mar-2000 #sequence\_revision 31-Mar-2000 #text\_change 09-Dec-2002  
 C/Accession: G81073  
 R/Petelin, H.; Saunders, N.J.; Heidelberg, J.; Jeffries, A.C.; Nelson, K.B.; Eisen, J.A  
 Hickey, E.K.; Haft, D.H.; Salzberg, S.L.; White, O.; Fleischmann, R.D.; Dougherty, B.A.;  
 Li, H.; Qin, H.; Yamashiro, J.; Gill, J.; Scarlato, V.; Mastignani, V.; Pizza, M.  
 Science 287, 1809-1815, 2000  
 A/Authors: Grandi, G.; Sun, L.; Smith, H.O.; Fraser, C.M.; Moxon, E.R.; Rappelli, R.; Ve  
 A/Title: Complete genome sequence of *Neisseria meningitidis* serogroup B strain MCS8.  
 A/Reference number: A81000; MUID:20175755; PMID:10710307  
 A/Accession: G81073  
 A/Status: preliminary  
 A/Molecule type: DNA  
 A/Residues: 1-160 <TET>  
 A/Cross-references: GB:AE002501; GB:AE002098; NID:G7226755; PIDN:AAF41868.1; PID:G722675  
 A/Experimental source: serogroup B, strain MCS8  
 C/Genetics:  
 A/Status: YACN  
 A/Reference number: 2C-methyl-D-erythritol 2,4-cyclodiphosphate synthase  
 Query Match 62.7%; Score 505; DB 2; Length 160;  
 Best Local Similarity 60.9%; Pred. No. 1.5e-39;  
 Matches 95; Conservative 21; Mismatches 40; Indels 0; Gaps 0;

QY 2 IRVKGKGVYHFNDDHIIIGGVKIPYKGLKSHSDGVVLAHALADALIGAAALGDIK 61  
 DB 4 IRVKGKGVYHFNDDHIIIGGVKIPYKGLKSHSDGVVLAHALADALIGAAALGDIK 63

QY 62 HEPDTPNFKGADSRVLLRHVYGIYKKGKLVNADVTIIAQPXKLPHVPGKRAIIAD 121  
 DB 62 HEPDTPNFKGADSRVLLRHVYGIYKKGKLVNADVTIIAQPXKLPHVPGKRAIIAD 123

QY 122 LETVDVFINVKATTEKLGFEGRKEGIAVQAVLIER 157  
 DB 122 LETVDVFINVKATTEKLGFEGRKEGIAVQAVLIER 159

RESULT 12  
 S70845  
 hypotetical protein 1 - *Haemophilus ducreyi*  
 C/Species: *Haemophilus ducreyi*  
 C/Date: 12-Feb-1998 #sequence\_revision 20-Feb-1998 #text\_change 09-Dec-2002  
 C/Accession: S70845  
 R/Palmer, K.L.; Munson Jr., R.S.  
 Mol. Microbiol. 18, 821-830, 1995  
 A/Title: Cloning and characterization of the genes encoding the haemolysin of *Haemophilu*  
 A/Reference number: S70843; MUID:96422469; PMID:8825086  
 A/Accession: S70845  
 A/Status: preliminary; nucleic acid sequence not shown; translation not shown  
 A/Molecule type: DNA  
 A/Residues: 1-160 <PAL>  
 A/Cross-references: EMBL:U02175; NID:G151070; PIDN:AA043539.1; PID:G151073  
 A/Note: the nucleotide sequence was submitted to the EMBL Data Library, July 1995  
 C/Suprafamily: 2C-methyl-D-erythritol 2,4-cyclodiphosphate synthase  
 Query Match 62.4%; Score 503; DB 2; Length 160;  
 Best Local Similarity 63.1%; Pred. No. 2.4e-39;  
 Matches 99; Conservative 17; Mismatches 41; Indels 0; Gaps 0;

QY 1 MIRVGKGVYHFNDDHIIIGGVKIPYKGLKSHSDGVVLAHALADALIGAAALGDIK 60  
 DB 1 MIRVGKGVYHFNDDHIIIGGVKIPYKGLKSHSDGVVLAHALADALIGAAALGDIK 60

QY 61 HEPDTPNFKGADSRVLLRHVYGIYKKGKLVNADVTIIAQPXKLPHVPGKRAIIAD 120  
 DB 61 HEPDTPNFKGADSRVLLRHVYGIYKKGKLVNADVTIIAQPXKLPHVPGKRAIIAD 120

Job time : 5.61946 secs

Search completed: January 29, 2004, 15:55:32  
Job time : 5.61946 secs

GenCore version 5.1.6  
Copyright (c) 1993 - 2004 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: January 29, 2004, 15:41:08 ; Search time 2.95645 Seconds  
(without alignments)  
2497.314 Million cell updates/sec

Title: US-09-941-947A-14  
Perfect score: 806  
Sequence: 1 MTRVGMGYDVHFRNDGDHII.....LGFGRKSGIAYQAVVIER 157

Scoring table: BLOSUM62  
Gapop 10.0 , Gapext 0.5

Searched: 127863 seqs, 47026705 residues  
Total number of hits satisfying chosen parameters: 127863

Minimum DB seq length: 0  
Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Database: SwissProt\_41.\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

## SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	575	71.3	157	1 ISPF_PSEAE	P57708 pseudomonas
2	554	68.7	158	1 ISPF_VIBCH	Q9KJ11 vibrio chol
3	551	68.4	158	1 ISPF_VIBVU	Q84C59 vibrio vuln
4	550	68.2	158	1 ISPF_VIBPA	Q87JQ3 vibrio para
5	542	67.2	162	1 ISPF_YERPE	Q8ZDP7 yerquina pe
6	537	66.6	159	1 ISPF_ECOL6	P36663 escherichia
7	537	66.6	159	1 ISPF_SALTI	Q8Z472 salmonella
8	537	66.6	159	1 ISPF_SALTY	Q8Z472 salmonella
9	537	66.6	159	1 ISPF_SALTY	Q8Z472 salmonella
10	534	66.3	158	1 ISPF_HAECU	P44615 haemophilus
11	529	65.6	158	1 ISPF_PASMU	P57954 pasteurella
12	520	64.5	157	1 ISPF_BACHD	Q9KJ17 bacillus ha
13	505	62.7	160	1 ISPF_NEIMB	Q9JYMS neisseria m
14	503	62.4	160	1 ISPF_HAEDU	Q47956 haemophilus
15	503	62.4	160	1 ISPF_NEIMA	Q9JYMS neisseria m
16	498	61.8	157	1 ISPF_CLOPE	Q8XJ08 clostridium
17	498	61.8	157	1 ISPF_CLOPE	Q8XJ08 clostridium
18	498	61.8	157	1 ISPF_CLOPE	Q8XJ08 clostridium
19	470	58.3	166	1 ISPF_XANAC	Q8XJ22 xanthomonas
20	456	56.6	163	1 ISPF_XANCP	Q8XJ20 xanthomonas
21	452	56.1	176	1 ISPF_XYLFA	Q9PCL5 xylella fas
22	451	56.0	157	1 ISPF_LISMO	Q8Y4B4 listeria mo
23	446	55.3	160	1 ISPF_FUSNN	Q8Z439 fusobacteri
24	443	55.0	157	1 ISPF_LISIN	Q8Z439 listeria in
25	429.5	53.3	160	1 ISPF_THERN	Q8Z439 thermocact
26	425	52.7	159	1 ISPF_CLOAB	Q8Z439 clostridium
27	424	52.6	155	1 ISPF_CLOAB	Q8Z439 clostridium
28	415	51.5	434	1 ISPF_RHIME	Q9Z490 r ispd/ispf
29	414	51.4	390	1 ISPF_BRIME	Q8Y4D8 b ispd/ispf
30	414	51.4	407	1 ISPF_RHILU	Q9Z490 r ispd/ispf
31	408	50.6	382	1 ISPF_CAUCR	Q9Z415 c ispd/ispf
32	403	50.0	161	1 ISPF_BUCAT	P57494 buchnera ap
33	403	50.0	165	1 ISPF_ANASP	Q8Y4F0 anabaena sp

34	403	50.0	167	1 ISPF_BUCAP	Q8X9D7 buchnera ap
35	403	50.0	371	1 ISPF_CAMTE	Q9PM68 c ispd/ispf
36	396	49.1	161	1 ISPF_SYNT3	P73426 synechocyst
37	383	47.5	236	1 ISPF_CATRO	Q9M4W3 catenanthu
38	373	46.3	231	1 ISPF_ARATH	Q9C4K8 arabidopsis
39	372	46.2	400	1 ISPF_AGRF5	Q8U4F4 a ispd/ispf
40	345	42.8	409	1 ISPF_HELPU	Q9ZML9 h ispd/ispf
41	343	42.6	406	1 ISPF_HELPU	Q25664 h ispd/ispf
42	329.5	40.9	187	1 ISPF_WIGGR	Q84224 wigglewort
43	329.5	40.9	187	1 ISPF_ZYMMO	Q9FZU1 z ispd/ispf
44	329	40.8	399	1 ISPF_TREPA	Q83525 t ispd/ispf
45	306.5	38.0	165	1 ISPF_THEMA	Q9W4B5 thermotoga

## ALIGNMENTS

RESULT 1  
ID ISPF\_PSEAE STANDARD; PRT: 157 AA.  
AC P57708; Q9HY03;  
DT 16-OCT-2001 (Rel. 40, last sequence update)  
DT 16-OCT-2001 (Rel. 40, last sequence update)  
DT 28-FEB-2003 (Rel. 41, last annotation update)  
DE 2-C-methyl-D-erythritol 2,4-cyclodiphosphate synthase (EC 4.6.1.12)  
DE (MECPs) (MECP-synthase).  
DE ISPF OR PA3627.  
OS Pseudomonas aeruginosa.  
OC Bacteria; Proteobacteria; Gammaproteobacteria; Pseudomonadales;  
OC Pseudomonadaceae; Pseudomonas.  
OX NCBI\_TaxID=287;  
RN (1)  
RP SEQUENCE FROM N.A.  
RC STRAIN=ATCC 15692 / PA01;  
RX MEDLINE=20437337; PubMed=10984043;  
RA Hickey M.U., Brinkman P.S.L., Hutmagle W.O., Kowalik D.J., Lagrou M.,  
RA Garber R.L., Goltz L., Tolentino E., Westbrock-Medman S., Yuan Y.,  
RA Brody L.L., Coulter S.N., Folger K.R., Kas A., Laidig K., Lim R.M.,  
RA Smith K.A., Spencer D.H., Wong G.K.-S., Wu Z., Paulsen I.T.,  
RA Feiler J., Sater M.H., Hancock R.E.W., Lory S., Olson M.V.,  
RT "Complete genome sequence of Pseudomonas aeruginosa PA01, an  
opportunistic pathogen".  
RT Nature 406:959-964(2000).  
RL Nature 406:959-964(2000).  
CC -1- FUNCTION: CONVERTS 4-DIPHOSPHOCYTIDYL-2C-METHYL-D-ERYTHRITOL 2-  
CC PHOSPHATE INTO 2C-METHYL-D-ERYTHRITOL 2,4-CYCLODIPHOSPHATE AND  
CC 2C-METHYL-D-ERYTHRITOL 3,4-CYCLODIPHOSPHATE AND CMP (BY SIMILARITY).  
CC -1- CATALYTIC ACTIVITY: 2-phospho-4-(cytidine 5'-diphospho)-2-C-  
CC methyl-D-erythritol = 2-C-methyl-D-erythritol 2,4-cyclodiphosphate  
CC + CMP.  
CC -1- PATHWAY: Nonaevonolone terpenoid biosynthesis pathway; fifth step.  
CC -1- SIMILARITY: BELONGS TO THE ISPF FAMILY.  
CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
CC between the Swiss Institute of Bioinformatics and the EMBL Outstation -  
CC the European Bioinformatics Institute. There are no restrictions on its  
CC use by non-profit institutions as long as its content is in no way  
CC modified and this statement is not removed. Usage by and for commercial  
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
CC  
CC EMBL: AE004782; AA007015.1;  
CC PIR: A83194; A83194.1;  
CC HAMAP: MF\_001077; -; 1.  
CC InterPro: IPR003526; YgBB.  
CC Pfam: PF02542; YgBB.1.  
CC TIGRFAMs: TIGR00151; ISPF.1.  
CC PROSITE: PS01350; ISPF.1.  
CC Lysase: Isoprene biosynthesis; Complete proteome.  
CC SEQUENCE 157 AA; 16662 MW; C39C377B6403BCA CMC64;  
Query Match 71.3%; Score 575; DB 1; Length 157;

Best Local Similarity 71.4%; Pred. No. 8,je-47;  
Matches 110; Conservative 17; Mismatches 27; Indels 0; Gaps 0;

QY 2 IRMGVGVYHRENDGDHIIIGVYKIPYKGLKHAHSDGVVLAHALADAILGAAALGDIGK 61  
DB 1 MIRIGGVYHRENDGDHIIIGVYKIPYKGLKHAHSDGVVLAHALADAILGAAALGDIGK 60  
QY 62 FPPTDPEFKGADSVLLRHVYGIYKGYLVNADVTIIAQAAPKAPHYVGRANITADL 121  
DB 61 FPPTDPEFKGADSVLLRHVYGIYKGYLVNADVTIIAQAAPKAPHYVGRANITADL 120  
QY 122 ETDVDFINVKATTTKXGFGKRGKGIAGVAVLIR 155  
DB 121 GVAVDQNVNKAATTTKXGFGKRGKGIAGVAVLIR 154

## RESULT 2

ISPF\_VIBCH STANDARD; PRT; 158 AA.

ID ISPF\_VIBCH STANDARD; PRT; 158 AA.  
AC 09KDU1;  
DT 16-OCT-2001 (Rel. 40, Last sequence update)  
DT 16-OCT-2001 (Rel. 40, Last sequence update)  
DT 28-FEB-2003 (Rel. 41, Last annotation update)  
DE 2-C-methyl-D-erythritol 2,4-cyclodiphosphate synthase (EC 4.6.1.12)  
DE (MECPS) (MECPS-synthase).  
ISPF OR V0529.  
GN Vibrio cholerae.  
OS Bacteria; Proteobacteria; Gammaproteobacteria; Vibrionales;  
OC Vibrionaceae; Vibrrio.  
OX NCBI\_TaxID=666;  
RN [1]  
RP SEQUENCE FROM N.A.  
RC STRAIN=E1 for N16961 / Serotype O1;  
RX MEDLINE=20406833; PubMed=10952301;  
RA Heidelberg J.F., Eisen J.A., Nelson W.C., Clayton R.A., Gwinn M.L.,  
RA Dodson R.J., Haft D.H., Hickey E.K., Peterson J.D., Umayam L.A.,  
RA Gill S.R., Nelson K.E., Read T.D., Tettelin H., Richardson D.,  
RA Ermolaeva K.D., Vamathevan J., Bass S., Qin H., Drygi I., Sellers P.,  
RA Salzberg S.L., Smith H.O., Colwell R.R., Mekalanos J.J., Venter J.C.,  
RA Fraser C.M.;  
RT "DNA sequence of both chromosomes of the cholera pathogen Vibrio  
cholerae";  
RL Nature 406:477-483(2000).  
CC -1- FUNCTION: CONVERTS 4-DIPHOSPHOCYTIDYL-2C-METHYL-D-ERYTHRITOL 2-  
PHOSPHATE INTO 2C-METHYL-D-ERYTHRITOL 2,4-CYCLODIPHOSPHATE AND  
CMP. ALSO CONVERTS 4-DIPHOSPHOCYTIDYL-2C-METHYL-D-ERYTHRITOL INTO  
2C-METHYL-D-ERYTHRITOL 3,4-CYCLOPHOSPHATE AND CMP (BY SIMILARITY).  
CC -1- CATALYTIC ACTIVITY: 2-phospho-4-(cytidine 5'-diphospho)-2-C-  
methyl-D-erythritol = 2-C-methyl-D-erythritol 2,4-cyclodiphosphate  
+ CMP.  
CC -1- PATHWAY: Nonaevulonate terpenoid biosynthesis pathway; fifth step.  
CC -1- SIMILARITY: BELONGS TO THE ISPF FAMILY.  
CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
between the Swiss Institute of Bioinformatics and the EMBL outstation -  
the European Bioinformatics Institute. There are no restrictions on its  
use by non-profit institutions as long as its content is in no way  
modified and this statement is not removed. Usage by and for commercial  
entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
CC EMBL: AB004139; AAF93697.1;  
CC PIR: D82311; D82311.  
CC TIGR: V0529;  
DR HAMAP: MF\_00107; -; 1.  
DR InterPro: IPR003526; YgDB.  
DR Pfam: PF02542; YgDB; 1.  
DR TIGRFAMs: TIGR00151; ispf; 1.  
DR PROSITE: PS01350; ISPF; 1.  
KM Lyase; Isoprene biosynthesis; Complete proteome.  
SQ SEQUENCE 158 AA; 16837 MW; 42B0AA40FEFFC6A CRC64;

Query Match 58.7%; Score 554; DB 1; Length 158;  
Best Local Similarity 70.7%; Pred. No. 7,6e-45;  
Matches 111; Conservative 16; Mismatches 30; Indels 0; Gaps 0;

QY 1 MIRVGGVYHRENDGDHIIIGVYKIPYKGLKHAHSDGVVLAHALADAILGAAALGDIGK 60  
DB 1 MIRIGGVYHRENDGDHIIIGVYKIPYKGLKHAHSDGVVLAHALADAILGAAALGDIGK 60  
QY 61 HPPDTPDFKAGDSVLLRHVYGIYKGYLVNADVTIIAQAAPKAPHYVGRANITADL 120  
DB 61 HPPDTPDFKAGDSVLLRHVYGIYKGYLVNADVTIIAQAAPKAPHYVGRANITADL 120  
QY 121 LETDVGFINVKATTTKXGFGKRGKGIAGVAVLIR 157  
DB 121 LETDVGFINVKATTTKXGFGKRGKGIAGVAVLIR 157

## RESULT 3

ISPF\_VIBVU STANDARD; PRT; 158 AA.

ID ISPF\_VIBVU STANDARD; PRT; 158 AA.  
AC 08DC59;  
DT 15-SEP-2003 (Rel. 42, Created)  
DT 15-SEP-2003 (Rel. 42, Last sequence update)  
DT 15-SEP-2003 (Rel. 42, Last annotation update)  
DE 2-C-methyl-D-erythritol 2,4-cyclodiphosphate synthase (EC 4.6.1.12)  
DE (MECPS) (MECPS-synthase).  
ISPF OR VY1583.  
GN Vibrio vulnificus.  
OS Bacteria; Proteobacteria; Gammaproteobacteria; Vibrionales;  
OC Vibrionaceae; Vibrrio.  
OX NCBI\_TaxID=672;  
RN [1]  
RP SEQUENCE FROM N.A.  
RC STRAIN=CWCP6;  
RX Rhee J.H., Kim S.Y., Chung S.S., Kim J.J., Moon Y.H., Jeong H.,  
RA Choy H.E.;  
RT "Complete genome sequence of Vibrio vulnificus CWCP6";  
RL Submitted (DEC-2002) to the EMBL/Genbank/DBJ databases.  
CC -1- FUNCTION: Converts 4-diphosphocytidylyl-2C-methyl-D-erythritol 2-  
phosphate into 2C-methyl-D-erythritol 2,4-cyclodiphosphate and  
CMP. Also converts 4-diphosphocytidylyl-2C-methyl-D-erythritol into  
2C-methyl-D-erythritol 3,4-cyclophosphate and CMP (By similarity).  
CC -1- CATALYTIC ACTIVITY: 2-phospho-4-(cytidine 5'-diphospho)-2-C-  
methyl-D-erythritol = 2-C-methyl-D-erythritol 2,4-cyclodiphosphate  
+ CMP.  
CC -1- PATHWAY: Nonaevulonate terpenoid biosynthesis pathway; fifth step.  
CC -1- SIMILARITY: BELONGS TO THE ISPF FAMILY.  
CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
between the Swiss Institute of Bioinformatics and the EMBL outstation -  
the European Bioinformatics Institute. There are no restrictions on its  
use by non-profit institutions as long as its content is in no way  
modified and this statement is not removed. Usage by and for commercial  
entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
CC EMBL: AB016802; AA010006.1;  
CC HAMAP: MF\_00107; -; 1.  
CC InterPro: IPR003526; YgDB.  
CC Pfam: PF02542; YgDB; 1.  
CC TIGRFAMs: TIGR00151; ispf; 1.  
CC PROSITE: PS01350; ISPF; 1.  
KM Isoprene biosynthesis; Lyase; Complete proteome.  
SQ SEQUENCE 158 AA; 17084 MW; 0E888F8050160112 CRC64;

Query Match 68.4%; Score 551; DB 1; Length 158;

Best Local Similarity 69.0%; Pred. No. 1.5e-44;  
Matches 107; Conservative 19; Mismatches 29; Indels 0; Gaps 0;

QY 1 MIRVGGVYHRENDGDHIIIGVYKIPYKGLKHAHSDGVVLAHALADAILGAAALGDIGK 60  
DB 1 MIRIGGVYHRENDGDHIIIGVYKIPYKGLKHAHSDGVVLAHALADAILGAAALGDIGK 60

QY 61 HPPDTPNFKGADSRVLLRHVYIVKEKGYKLVNADVTIIAQAQKMLPHYEGKANIAD 120  
 DB 61 HPPDTPNFKGADSRVLLRHVYIVKEKGYKLVNADVTIIAQAQKMLPHYEGKANIAD 120  
 QY 121 LETDVPFINVATTEKLGFEGRKEGIAVQAVLI 155  
 DB 121 LETDVPFINVATTEKLGFEGRKEGIAVQAVLI 155

## RESULT 4

ISPF\_VIBPA STANDARD; PRT; 158 AA.  
 ID 087103;  
 DT 15-SEP-2003 (Rel. 42, Created)  
 DT 15-SEP-2003 (Rel. 42, Last sequence update)  
 DT 15-SEP-2003 (Rel. 42, Last annotation update)  
 DE 2-C-methyl-D-erythritol 2,4-cyclodiphosphate synthase (EC 4.6.1.12)  
 DE (MECPs) (MECDP-synthase).  
 GN ISPF OR VP2558.  
 OS Vibrio parahaemolyticus.  
 OC Bacteria; Proteobacteria; Gammaproteobacteria; Vibrionales;  
 OC Vibrionaceae; Vibrio.  
 CC NCBI\_TaxID=670;  
 RX MEDLINE=22508454; PubMed=12620739;  
 RA Makino K., Oshima K., Kurokawa K., Yokoyama K., Uda T., Tagomori K.,  
 RA Iijima Y., Nishina M., Nakano M., Yamashita A., Kubota Y., Kimura S.,  
 RA Yasunaga T., Honda T., Shinagawa H., Hattori M., Iida T.,  
 RT Genome sequence of Vibrio parahaemolyticus: a pathogenic mechanism  
 RT distinct from that of V. cholerae.";  
 RL Lancel 361:743-749(2003).

-1- FUNCTION: Converts 4-diphosphocytidylyl-2C-methyl-D-erythritol 2-  
 phosphate into 2C-methyl-D-erythritol 2,4-cyclodiphosphate and  
 CMP. Also converts 4-diphosphocytidylyl-2C-methyl-D-erythritol into  
 2C-methyl-D-erythritol 3,4-cyclodiphosphate and CMP (by similarity).  
 CC CATALYTIC ACTIVITY: 2-phospho-4-(cytidine 5'-diphospho)-2-C-  
 methyl-D-erythritol = 2-C-methyl-D-erythritol 2,4-cyclodiphosphate  
 + CMP.  
 CC -1- PATHWAY: Nonaevionate terpenoid biosynthesis pathway; fifth step.  
 CC -1- SIMILARITY: BELONGS TO THE ISPF FAMILY.

CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
 CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).

DR EMBL; AP005081; BAC60821.1; -  
 DR HAMAP; MF 00107; -; 1.  
 DR PROSITE; PS01350; ISPF, 1.  
 KW Isoprene biosynthesis; lyase; Complete proteome.  
 SQ SEQUENCE 158 AA; 16949 MW; F8C9FPAE03D3D1 CRC64;

Query Match 68.2%; Score 550; DB 1; Length 158;

Best Local Similarity 69.0%; Pred. No. 1.8e-44;  
 Matches 107; Conservative 19; Mismatches 29; Indels 0; Gaps 0;

QY 1 MIRVGAGYVHRFNDGHIILGVKIPYRKGLAHSDDGVVLAALDAILGAALGDICK 60  
 DB 1 MIRVGAGYVHRFNDGHIILGVKIPYRKGLAHSDDGVVLAALDAILGAALGDICK 60  
 QY 1 MINGHGFVHAKGSGEPVILIGVAFPEOGLAHSDDGVVLAALDAILGAALGDICK 60  
 DB 1 MINGHGFVHAKGSGEPVILIGVAFPEOGLAHSDDGVVLAALDAILGAALGDICK 60  
 QY 61 HPPDTPNFKGADSRVLLRHVYIVKEKGYKLVNADVTIIAQAQKMLPHYEGKANIAD 120  
 DB 61 HPPDTPNFKGADSRVLLRHVYIVKEKGYKLVNADVTIIAQAQKMLPHYEGKANIAD 120  
 QY 121 LETDVPFINVATTEKLGFEGRKEGIAVQAVLI 155  
 DB 121 LETDVPFINVATTEKLGFEGRKEGIAVQAVLI 155  
 QY 121 LETDVPFINVATTEKLGFEGRKEGIAVQAVLI 155  
 DB 121 LETDVPFINVATTEKLGFEGRKEGIAVQAVLI 155

RESULT 5  
 ID ISPF\_YERPE STANDARD; PRT; 162 AA.  
 AC 062BP7;  
 DT 28-FEB-2003 (Rel. 41, Created)  
 DT 28-FEB-2003 (Rel. 41, Last sequence update)  
 DT 28-FEB-2003 (Rel. 41, Last annotation update)  
 DE 2-C-methyl-D-erythritol 2,4-cyclodiphosphate synthase (EC 4.6.1.12)  
 DE (MECPs) (MECDP-synthase).  
 GN ISPF OR YPO3360 OR Y0829.  
 OS Yersinia pestis.  
 OC Bacteria; Proteobacteria; Gammaproteobacteria; Enterobacteriales;  
 OC Enterobacteriaceae; Yersinia.  
 CC NCBI\_TaxID=632;  
 RX MEDLINE=21470413; PubMed=11586360;  
 RA Parkhill J., Wren B.W., Thomson N.R., Tibball R.W., Holden M.T.G.,  
 RA Prentice M.B., Sebahia M., James K.D., Churcher C., Mungall K.L.,  
 RA Baker S., Basham D., Bentley S.D., Brooks K., Cerdeno-Tarraga A.M.,  
 RA Chillingworth T., Cronin A., Davies R.M., Davis P., Dougan G.,  
 RA Felwell T., Hamlin N., Holroyd S., Jagsels K., Karlyshev A.V.,  
 RA Leather S., Moule S., Oyston P.C.F., Quail M., Rutherford K.,  
 RA Simmonds M., Skelton J., Stevens K., Whitehead S., Barrall B.G.,  
 RT Genome sequence of Yersinia pestis, the causative agent of plague.";  
 RL Nature 413:523-527(2001).

-1- FUNCTION: Converts 4-diphosphocytidylyl-2C-methyl-D-erythritol 2-  
 phosphate into 2C-methyl-D-erythritol 2,4-cyclodiphosphate and  
 CMP. Also converts 4-diphosphocytidylyl-2C-methyl-D-erythritol into  
 2C-methyl-D-erythritol 3,4-cyclodiphosphate and CMP (by similarity).  
 CC CATALYTIC ACTIVITY: 2-phospho-4-(cytidine 5'-diphospho)-2-C-  
 methyl-D-erythritol = 2-C-methyl-D-erythritol 2,4-cyclodiphosphate  
 + CMP.  
 CC -1- PATHWAY: Nonaevionate terpenoid biosynthesis pathway; fifth step.  
 CC -1- SIMILARITY: BELONGS TO THE ISPF FAMILY.

CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
 CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).

DR EMBL; AJ414156; CAC92590.1; -  
 DR EMBL; AE013685; AAM84414.1; ALT\_INIT.  
 DR PIR; AB0408; AB0408.  
 DR HAMAP; MF 00107; -; 1.  
 DR InterPro; IPR003526; YGPB.  
 DR Pfam; PF02542; YGPB; 1.  
 DR TIGRPFAM; TIGR00151; ISPF, 1.  
 DR PROSITE; PS01350; ISPF, 1.  
 KW Lyase; isoprene biosynthesis; Complete proteome.  
 SQ SEQUENCE 162 AA; 17182 MW; 9A295411D3DF1EB CRC64;

Query Match 67.2%; Score 542; DB 1; Length 162;  
 Best Local Similarity 68.6%; Pred. No. 1e-43;  
 Matches 107; Conservative 17; Mismatches 30; Indels 2; Gaps 1;

QY 2 IRVAGYVHRF--NDGHIILGVKIPYRKGLAHSDDGVVLAALDAILGAALGDIG 59  
 DB 2 IRVAGYVHRF--NDGHIILGVKIPYRKGLAHSDDGVVLAALDAILGAALGDIG 59

Db 1 MRIGGVDVHFGSNGSGPLIGGVR:PYEKGGLAHSDGDVALHAADALGAAAGDIG 60  
 Qy 60 KHPDPTDPPNFGASRVLLRHVYGVYKEKGYKLVNADVTITIAQAPKPLPHVPGKRAITIA 119  
 Db 61 KLPFDPTDPAFAGASRGRLRRAVYRIIAKGYKLNLDITITIAQAPKRAPIHPCQKRVNLAE 120  
 Qy 120 DLRTDVPDINKATTTTEKLGPEGRKEGIAVOAVL 155  
 Db 121 DLQCHMDINKATTTTDLGFTGEGIGIACRAVALL 156

RESULT 6  
 ISPF\_ECOL6  
 ID ISPF\_SCOL6 STANDARD; PRT; 159 AA.

AC Q8REJ6;  
 DT 15-SEP-2003 (Rel. 42, Created)  
 DT 15-SEP-2003 (Rel. 42, Last sequence update)  
 DT 15-SEP-2003 (Rel. 42, Last annotation update)  
 DB 2-C-methyl-D-erythritol 2,4-cyclodiphosphate synthase (EC 4.6.1.12)  
 DB (MECPs) (MECDP-synthase).  
 DB ISPF OR C313.

OS Escherichia coli 06  
 OC Bacteria; Proteobacteria; Gammaproteobacteria; Enterobacteriales;  
 OC Enterobacteriaceae; Escherichia.  
 OX NCBI\_TaxID=217992;

SEQUENCE FROM N.A.  
 RP STRAIN=O6:H1 / ATCC 700928;  
 RC MEDLINE=2386234; PubMed=12471157;

RA Raeko R.A., Burland V., Plunkett G. III, Redford P., Roesch P.,  
 RA Raeko D., Buckles E.L., Lau S.-R., Boutin A., Hackett J., Stroud D.,  
 RA Mayhew G.F., Rose D.J., Zhou S., Schwartz D.C., Perna N.T.,  
 RA Mobley H.L.T., Donnenberg M.S., Blattner F.R.;  
 RT "Extensive mosaic structure revealed by the complete genome sequence  
 of uropathogenic Escherichia coli";

Proc. Natl. Acad. Sci. U.S.A. 99:17020-17024(2002).

CC -1- FUNCTION: Converts 4-diphosphocytidyl-2C-methyl-D-erythritol 2-  
 CC phosphate into 2C-methyl-D-erythritol 2,4-cyclodiphosphate and  
 CC CMP. Also converts 4-diphosphocytidyl-2C-methyl-D-erythritol into  
 CC 2C-methyl-D-erythritol 3,4-cyclodiphosphate and CMP (by similarity).  
 CC -1- CATALYTIC ACTIVITY: 2-phospho-4-(cytidine 5'-diphospho)-2-C-  
 CC methyl-D-erythritol = 2-C-methyl-D-erythritol 2,4-cyclodiphosphate  
 CC + CMP.

CC -1- PATHWAY: Nonaevulonate terpenoid biosynthesis pathway; fifth step.  
 CC -1- SIMILARITY: BELONGS TO THE ISPF FAMILY.

CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
 CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).

CC EMBL: AE016765; AAN81762.1;  
 CC HAMAP: MF 00107; -1.1;  
 CC InterPro: IPR003526; Igb3.

DR InterPro: IPR003526; Igb3.  
 DR Pfam: PF02542; Igb3; 1.  
 DR TIGRFAMs: TIGR00151; ISPF; 1.  
 DR PROSITE: PS01350; ISPF; 1.

KV Isoprene biosynthesis; lyase; Complete proteome.  
 RN Isoprene biosynthesis; lyase; Complete proteome.  
 SO SEQUENCE 159 AA; 16867 MM; 8PD4573623BA4549 CRC64;

Query Match 66.6%; Score 537; DB 1; Length 159;  
 Best Local Similarity 69.5%; Pred. No. 3e-43;

Matches 107; Conservative 14; Mismatches 33; Indels 0; Gaps 0;

Qy 2 IRVAGYDVHFGSNGSGPLIGGVR:PYEKGGLAHSDGDVALHAADALGAAAGDIG 61  
 Db 1 MRIGGVDVHFGSNGSGPLIGGVR:PYEKGGLAHSDGDVALHAADALGAAAGDIG 60  
 Qy 62 FPDPTDPPNFGASRVLLRHVYGVYKEKGYKLVNADVTITIAQAPKPLPHVPGKRAITIA 121

Db 61 FPDPTDPPNFGASRVLLRHVYGVYKEKGYKLVNADVTITIAQAPKPLPHVPGKRAITIA 120  
 Qy 122 FPDPTDPPNFGASRVLLRHVYGVYKEKGYKLVNADVTITIAQAPKPLPHVPGKRAITIA 155  
 Db 121 GCHMDVNVKATTTTEKLGPEGRKEGIAVOAVL 154

RESULT 7

ISPF\_ECOL1  
 ID ISPF\_SCOL1 STANDARD; PRT; 159 AA.

AC P36663;  
 DT 01-JUN-1994 (Rel. 29, Created)  
 DT 01-JUN-1994 (Rel. 29, Last sequence update)  
 DT 15-SEP-2003 (Rel. 42, Last annotation update)  
 DB 2-C-methyl-D-erythritol 2,4-cyclodiphosphate synthase (EC 4.6.1.12)  
 DB (MECPs) (MECDP-synthase).  
 DB ISPF OR MECS OR B2746 OR Z4054 OR EC53600 OR SF2769.  
 OS Escherichia coli  
 OS Escherichia coli 0157:H7, and  
 OS Shigella flexneri.

OC Bacteria; Proteobacteria; Gammaproteobacteria; Enterobacteriales;  
 OC Enterobacteriaceae; Escherichia.  
 OX NCBI\_TaxID=562; 83334; 623;

SEQUENCE FROM N.A.

RP SPECIES=E.coli; STRAIN=MP180;  
 RC MEDLINE=95014035; PubMed=7928962;

RA U.C., Ichikawa J.K., Ravetto J.J., Kuo H.-C., Fu J.C., Clarke S.;  
 RT "A new gene involved in stationary-phase survival located at 59  
 RT minutes on the Escherichia coli chromosome";  
 RT J. Bacteriol. 176:6015-6022(1994).

RP SEQUENCE FROM N.A., PARTIAL SEQUENCE, AND CHARACTERIZATION.

RC SPECIES=E.coli;  
 RX MEDLINE=20183914; PubMed=10694574;

RA Herz S., Mungstentweck J., Schuhr C.A., Hecht S., Luettgen H.,  
 RA Sagner S., Fellemeier W., Bisenreich W., Zenk M.H., Bachner A.,  
 RA Rohdich F.;  
 RT "Biosynthesis of terpenoids: YgbB protein converts 4-diphosphocytidyl-  
 RT 2C-methyl-D-erythritol 2-phosphate to 2C-methyl-D-erythritol 2,4-  
 RT cyclodiphosphate";  
 RT Proc. Natl. Acad. Sci. U.S.A. 97:2486-2490(2000).

CC [3] SEQUENCE FROM N.A., AND CHARACTERIZATION.

CC SPECIES=E.coli;  
 RC TAKEGI M., Kuzuyama T., Kaneda K., Watanabe H., Dairi T., Seto H.;

RT "Studies on the nonmevalonate pathway: formation of 2-C-methyl-D-  
 RT erythritol 2,4-cyclodiphosphate from 2-phospho-4-(cytidine 5'-  
 RT diphospho)-2-C-methyl-D-erythritol";  
 RT Tetrahedron Lett. 41:3395-3398(2000).

CC [4] SEQUENCE FROM N.A.

CC SPECIES=E.coli; STRAIN=X12 / MG1655;  
 RX MEDLINE=97426617; PubMed=9278503;

RA Blattner F.R., Plunkett G. III, Bloch C.A., Perna N.T., Burland V.,  
 RA Riley M., Collado-Vides J., Glasner J.D., Rode C.K., Mayhew G.F.,  
 RA Gregor J., Davis N.W., Kirkpatrick H.A., Goeden M.A., Rose D.J.,  
 RA Mau B., Shao Y.;

RT "The complete genome sequence of Escherichia coli K-12";  
 RT Science 277:1453-1474(1997).

CC [5] SEQUENCE FROM N.A.

CC SPECIES=E.coli; STRAIN=O157:H7 / EDL933 / ATCC 700927;  
 RX MEDLINE=21074935; PubMed=11206551;

RA Perna N.T., Plunkett G. III, Burland V., Mau B., Glasner J.D.,  
 RA Rose D.J., Mayhew G.F., Evans P.S., Gregor J., Kirkpatrick H.A.,  
 RA Joshi G., Hackett J., Klink S., Boutin A., Shao Y., Miller L.,  
 RA Grobeck S.J., Davis N.W., Lim A., Dimalanta E.T., Potamousis K.,  
 RA Apodoca J., Annatharaman T.S., Lin J., Yen G., Schwartz D.C.,  
 RA Welch R.A., Blattner F.R.;

RT "Genome sequence of enterohaemorrhagic Escherichia coli O157:H7";  
 RT Nature 409:529-533(2001).

**S0** SEQUENCE 159 AA; 16897 MW; 9FCS563623A62939 CRC64;

Query Match 66.6%; Score 537; DB 1; Length 159;  
Best Local Similarity 69.5%; Pred. No. 3e-43;  
Matches 107; Conservative 14; Mismatches 33; Indels 0; Gaps 0

**Dy** 2 IRVGMQVDYHRRNDNDHHIIIGCVKI PYRKGLEASDGGVVTALADALIGAAAGDGK 61  
:::|||||:::|||||:::|||||:::|||||:::|||||:::|||||:::|||||:::|||||  
**Db** 1 MRGHSFDVAHQSGRGPIIIGCVRI PRYKGLANSDGVALLHTDLTGAAALGDIGKL 60  
:::|||||:::|||||:::|||||:::|||||:::|||||:::|||||:::|||||:::|||||

**Dy** 62 FPDPTDNFNFGADSRYLLHRHYIVKESEKYKLVNADVITIAQAPEMLPHVPGSRANIADL 121  
:::|||||:::|||||:::|||||:::|||||:::|||||:::|||||:::|||||:::|||||  
**Db** 61 FPDTDAFAFGADSRRLHEAMRIQAKYTILGANDVTIILAEPGSLPHIQVEVFIAEDL 120  
:::|||||:::|||||:::|||||:::|||||:::|||||:::|||||:::|||||:::|||||

**Dy** 122 EIDVDPIFNKATTEKTGEGEKGAGIAOVAVL 155  
:::|||||:::|||||:::|||||:::|||||:::|||||:::|||||:::|||||:::|||||  
**Db** 121 GCAMDVNFKATTEKTGTGTGGBGIACEAVALL 154  
:::|||||:::|||||:::|||||:::|||||:::|||||:::|||||:::|||||:::|||||

**RESULT 8**

**ISPF\_SALT**

ID ISPF\_SALT STANDARD; PRT; 159 AA.

AC Q82472;  
DT 28-FEB-2003 [Rel. 41, Created]  
DT 28-FEB-2003 [Rel. 41, Last sequence update]  
DT 15-SEP-2003 [Rel. 42, Last annotation update]  
DE 2-C-methyl-D-erythritol 2,4-cyclodiphosphate synthase (EC 4.6.1.12)  
GN (MECS) (MCDP-synthase).  
NP ISPF OR STY3054 OR T2830.  
OS Salmomella typhi.  
OC Bacteria; Proteobacteria; Gammaproteobacteria; Enterobacteriales;  
OC Enterobacteriaceae; Salmomella.  
OX NCBI\_TaxId=601.  
RN [1]  
RP SEQUENCE FROM N.A.  
RC STRAIN=CT18;  
RX MEDLINE=21534947; PubMed=11677608;  
RA Parikhil J., Dougan G., James K.D., Thomson N.R., Pickard D., Wain J., Churcher C., Mungall K.L., Bentley S.D., Holden M.T.G., Sebatina M., Baker S., Basham D., Brooks K., Chillingworth T., Connor P., Cronin A., Davis P., Davies R.M., Dowd L., White N., Farrar J., Felwell T., Hamlin N., Haque A., Hien T.T., Holtroyd S., Jagels K., Krogh A., Larsen T.S., Leachter S., Moule S., O'Gaora P., Parry C., Quail M., Rutherford K., Simmonds M., Skelton J., Stevens K., Whitehead S., Barrett B.G.;  
RT "Complete genome sequence of a multiple drug resistant Salmomella enterica serovar Typhi CT18.";  
RV Nature 413:848-852(2001).  
RW [2]

RN SEQUENCE FROM N.A.  
RC STRAIN=Ty2 / ATCC 700931;  
RX MEDLINE=22531367; PubMed=12644504;  
RA Deng M., Ikon S.-R., Plunkett G., III, Mayhew G.F., Rose D.J., Harland V., Kodoyanni V., Schwartz D.C., Blatter F.R.;  
RT "Comparative genomes of Salmomella enterica serovar Typhi strains Ty2 and CT18.";  
RV J. Bacteriol. 185:2330-2337(2003).

-1- FUNCTION: Inverts 4-diphosphocytidyl-2C-methyl-D-erythritol 2-phosphate into 2C-methyl-D-erythritol 2,4-cyclodiphosphate and CMP. Also converts 4-diphosphocytidyl-2C-methyl-D-erythritol into 2C-methyl-D-erythritol 3,4-cyclophosphate and CMP (by similarity).  
-1- CATALYTIC ACTIVITY: 2-phospho-4-(cytidine 5'-diphospho)-2-C-methyl-D-erythritol = 2-C-methyl-D-erythritol 2,4-cyclodiphosphate + CMP.  
-1- PATHWAY: Nonmevalonate terpenoid biosynthesis pathway; fifth step.  
-1- SIMILARITY: BELONGS TO THE ISPF FAMILY.

-----  
THIS SWISS-PROT entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL outstation at the European Bioinformatics Institute. There are no restrictions on its use by non-profit institutions as long as its content is in no way distorted or modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (See <http://www.isb-sib.ch/announce/>)

CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch).  
 CC EMBL: AL627276; CAD06035.1; -  
 DR STGENE: SG27272; ISPF.  
 DR HAMAP: MF\_00107; -; 1  
 DR InterPro: IPR003526; Ygdb.  
 DR Pfam: PF02542; Ygdb; 1.  
 DR TIGRfam: TIGR00151; ISPF; 1.  
 DR PROSITE: PS01350; ISPF; 1.  
 KM Lyase; Isoprene biosynthesis; Complete proteome.  
 SQ SEQUENCE 159 AA; 16885 MW; 9FD9263625662939 CRC64;

Query Match 66.6%; Score 537; DB 1; Length 159;  
 Best Local Similarity 69.5%; Pred. No. 3e-43;  
 Matches 107; Conservative 14; Mismatches 33; Indels 0; Gaps 0;  
 QY 2 IRVGMGVVHREDFNDGHIILGKVKIPYKGLKSHSDGCVLHATADATIGAAAGDICK 61  
 DB 1 MRIGGFVHAFGSGPFIIGVRIPEKGLKSHSDGCVLHATADATIGAAAGDICK 60  
 QY 62 PPDTDPNFKGADSRVLLRHVYGIYKKGKYLWNADVTIIAQAAPKMLPHVPGKRAIAADL 121  
 DB 61 PPDTDPNFKGADSRVLLRHVYGIYKKGKYLWNADVTIIAQAAPKMLPHVPGKRAIAADL 120  
 QY 122 ETDVDFINVKATTEKLGFEKRGKGIAGVAVLI 155  
 DB 121 GCHMDVNVKATTEKLGFEKRGKGIAGVAVLI 154

RESULT 9  
 ISPF SALTY STANDARD; PRT; 159 AA.  
 AC Q6ZNF7;  
 DT 28-FEB-2003 (Rel. 41, Created)  
 DT 28-FEB-2003 (Rel. 41, Last sequence update)  
 DT 28-FEB-2003 (Rel. 41, Last annotation update)  
 DE 2-C-methyl-D-erythritol 2,4-cyclodiphosphate synthase (EC 4.6.1.12)  
 GN (MECPD) (MECPD-synthase).  
 OS ISPF OR STM2929.  
 OC Salmonella typhimurium.  
 OC Bacteria; Proteobacteria; Gammaproteobacteria; Enterobacteriales;  
 OC Enterobacteriaceae; Salmonella.  
 OX NCBI\_TaxID=602;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=LT2 / SSGC1412 / AFCC 700720;  
 RX MEDLINE=2154948; PubMed=11677609;  
 RA McLeiland M., Sanderson K.E., Spieth J., Clifton S.W., Latreille P.,  
 RA Courtney L., Portwolk S., Ali J., Dante M., Du P., Hou S., Layman D.,  
 RA Leonard S., Nguyen C., Scott K., Holmes A., Grewal N., Mulvaney B.,  
 RA Ryan E., Sun H., Florin L., Miller W., Stoneking T., Nhan M.,  
 RA Waterston R., Wilson R.K.;  
 RT "Complete genome sequence of *Salmonella enterica* serovar Typhimurium  
 LT2".  
 RL Nature 413:852-856(2001).  
 CC -1- FUNCTION: Converts 4-diphosphocytidyl-2C-methyl-D-erythritol 2-  
 CC phosphate into 2C-methyl-D-erythritol 2,4-cyclodiphosphate and  
 CC CMP. Also converts 4-diphosphocytidyl-2C-methyl-D-erythritol into  
 CC 2C-methyl-D-erythritol 3,4-cyclophosphate and CMP (By similarity).  
 CC -1- CATALYTIC ACTIVITY: 2-phospho-4-(cytidine 5'-diphospho)-2-C-  
 CC methyl-D-erythritol = 2-C-methyl-D-erythritol 2,4-cyclodiphosphate  
 CC + CMP.  
 CC -1- PATHWAY: Nonmevalonate terpenoid biosynthesis pathway; fifth step.  
 CC -1- SIMILARITY: BELONGS TO THE ISPF FAMILY.

CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
 CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).

DR EMBL: AB008833; AAL21809.1; -  
 DR STGENE: SG27272; ISPF.  
 DR HAMAP: MF\_00107; -; 1  
 DR InterPro: IPR003526; Ygdb.  
 DR Pfam: PF02542; Ygdb; 1.  
 DR TIGRfam: TIGR00151; ISPF; 1.  
 DR PROSITE: PS01350; ISPF; 1.  
 KM Lyase; Isoprene biosynthesis; Complete proteome.  
 SQ SEQUENCE 159 AA; 16899 MW; 9FD88D2C95662939 CRC64;

Query Match 66.6%; Score 537; DB 1; Length 159;  
 Best Local Similarity 69.5%; Pred. No. 3e-43;  
 Matches 107; Conservative 14; Mismatches 33; Indels 0; Gaps 0;  
 QY 2 IRVGMGVVHREDFNDGHIILGKVKIPYKGLKSHSDGCVLHATADATIGAAAGDICK 61  
 DB 1 MRIGGFVHAFGSGPFIIGVRIPEKGLKSHSDGCVLHATADATIGAAAGDICK 60  
 QY 62 PPDTDPNFKGADSRVLLRHVYGIYKKGKYLWNADVTIIAQAAPKMLPHVPGKRAIAADL 121  
 DB 61 PPDTDPNFKGADSRVLLRHVYGIYKKGKYLWNADVTIIAQAAPKMLPHVPGKRAIAADL 120  
 QY 122 ETDVDFINVKATTEKLGFEKRGKGIAGVAVLI 155  
 DB 121 GCHMDVNVKATTEKLGFEKRGKGIAGVAVLI 154

RESULT 10  
 ISPF BACSU STANDARD; PRT; 158 AA.  
 AC Q06756;  
 DT 01-JUN-1994 (Rel. 29, Created)  
 DT 01-JUN-1994 (Rel. 29, Last sequence update)  
 DT 28-FEB-2003 (Rel. 41, Last annotation update)  
 DE 2-C-methyl-D-erythritol 2,4-cyclodiphosphate synthase (EC 4.6.1.12)  
 GN (MECPD) (MECPD-synthase).  
 OS Bacillus subtilis.  
 OC Bacteria; Firmicutes; Bacillales; Bacillaceae; Bacillus.  
 OX NCBI\_TaxID=1423;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC MEDLINE=94171772; PubMed=7510287;  
 RX Borsari R., Boretto R., Putzer H., Pelchat M., Grunberg-Manago M.,  
 RA Lepointe J.;  
 RT "Clustering and co-transcription of the *Bacillus subtilis* genes  
 RT encoding the aminoacyl-tRNA synthetases specific for glutamate and for  
 RT cysteine and the first enzyme for cysteine biosynthesis".  
 RL J. Biol. Chem. 269:7473-7482(1994).  
 RN [2]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=168;  
 RX MEDLINE=96051385; PubMed=7584024;  
 RA Ogasawara N., Nakai S., Yoshikawa H.;  
 RT "Systematic sequencing of the 180 kilobase region of the *Bacillus*  
 RT subtilis chromosome containing the replication origin".  
 RL DNA Res. 1:1-14(1994).  
 RN [3]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=168;  
 RX MEDLINE=98044033; PubMed=9384377;  
 RA Kunitz F., Ogasawara N., Moser I., Albertini A.M., Alloni G.,  
 RA Azevedo V., Boretto M.G., Bessieres P., Bolotin A., Borchert S.,  
 RA Borsari R., Boretto L., Brans A., Braun M., Brigelli S.C., Bron S.,  
 RA Brouillet S., Brusch C.V., Caldwell B., Capuano V., Carter N.M.,  
 RA Cho S.K., Codani J.J., Comerton I.F., Cummings N.J., Daniel R.A.,  
 RA Denzot F., Devine K.M., Dusterhoft A., Ehrlich S.D., Emerson P.T.,  
 RA Eutlian K.D., Erington J., Fabret C., Ferrati E., Foulger D.,  
 RA Fritz C., Fujita M., Fujita Y., Fuma S., Galizzi A., Galleron N.,  
 RA Ghim S.Y., Glaser P., Goffeau A., Golightly E.J., Grandi G.,  
 RA Giuseppe G., Guy B.J., Haga K., Halech J., Harwood C.R., Henaut A.,  
 RA Hilbert H., Holteppel S., Hosono S., Hullo M.F., Itaya M., Jones L.,  
 RA Joris B., Karamata D., Kaishara Y., Klaerr-Blanchard M., Klein C.,

RA Kobayashi Y., Koelter P., Koningsstein G., Krogh S., Kumano M.,  
 RA Kunita K., Lapius A., Lardinois S., Lauber J., Lazarevic V.,  
 RA Lee S.M., Levine A., Liu H., Masuda S., Manuel C., Medigue C.,  
 RA Medina N., Mellado R., Mizuno M., Moestl D., Nakai S., Nodack M.,  
 RA Noone D., O'Reilly M., Ogawa K., Ogihara A., Oudaga B., Park S.H.,  
 RA Paro V., Pohl T.M., Portetalle D., Portwillek S., Prescott A.M.,  
 RA Presecan E., Pujic P., Purnelle B., Rapoport G., Rey M., Reynolds S.,  
 RA Rieger M., Rivolta C., Rocha E., Roche B., Rose M., Sadate Y.,  
 RA Sato T., Scanlan E., Schleich S., Schroeter R., Scoffone F.,  
 RA Sekiguchi J., Sekowska A., Seror S.J., Seror P., Shin B.S., Soldo B.,  
 RA Sorokin A., Tacconi E., Takagi T., Takahashi H., Takemaru K.,  
 RA Takeuchi M., Tanakoshi A., Tanaka T., Terstra P., Tognoni A.,  
 RA Tosato V., Uchiyama S., Vandenbol M., Vamler F., Vassarotti A.,  
 RA Vairi A., Wambutt R., Wedler E., Wedler H., Wetzengger T.,  
 RA Winters P., Wiput A., Yamamoto H., Yamane K., Yasumoto K., Yata K.,  
 RA Yoshida K., Yoshikawa H.F., Zumbstein E., Yoshikawa H., Zundin A.,  
 RT "The complete genome sequence of the Gram-positive bacterium *Bacillus*  
 RT *subtilis*,"  
 RU Nature 350:249-256(1997).  
 CC -1- FUNCTION: CONVERTS 4-DIPHOSPHOCYTIDYL-2C-METHYL-D-ERYTHRITOL 2-  
 CC PHOSPHATE INTO 2C-METHYL-D-ERYTHRITOL 2,4-CYCLODIPHOSPHATE AND  
 CC CMP. ALSO CONVERTS 4-DIPHOSPHOCYTIDYL-2C-METHYL-D-ERYTHRITOL INTO  
 CC 2C-METHYL-D-ERYTHRITOL 3,4-CYCLOPHOSPHATE AND CMP (BY SIMILARITY).  
 CC -1- CATALYTIC ACTIVITY: 2-phospho-4-(cytidine 5'-diphospho)-2-C-  
 CC methyl-D-erythritol = 2-C-methyl-D-erythritol 2,4-cyclodiphosphate  
 CC + CMP.  
 CC -1- PATHWAY: Nonaevonate terpenoid biosynthesis pathway; fifth step.  
 CC -1- SIMILARITY: BELONGS TO THE ISPF FAMILY.  
 CC -----  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation-  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
 CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
 CC -----  
 CC EMBL, U4589; AAA21795.1; --  
 CC EMBL, D6185; BAA05325.1; --  
 CC EMBL, Z99104; CAB11867.1; --  
 CC PIR, F69741; F69741.  
 CC Subtilist; BG1053; ispf.  
 CC HAMAP; MF 00107; -; 1.  
 CC InterPro; IPR003526; Ygbb.  
 CC Pfam; PF02542; Ygbb; 1.  
 CC TIGRFAWS; TIGR00151; ispf; 1.  
 CC PROSITE; PS01350; ispf; 1.  
 CC Lyase; Isoprene biosynthesis; Complete proteome.  
 CC SEQUENCE 158 AA; 17126 MW; ED11D03BC752BD0 CRC64;  
 SQ  
 Query Match 66.3%; Score 534; DB 1; Length 158;  
 Best Local Similarity 65.6%; Pred. No. 5, 6e-43;  
 Matches 103; Conservative 18; Mismatches 36; Indels 0; Gaps 0;  
 QY 1 MIRVGMYDVARENDGHIILIGVKIPYKGLAHSDDGVVLAALADALIGAAALGDIK 60  
 DB 1 MERIGGFVHGVHVRPLILIGIEIPIYKGLGHSDAVVLAHTVADALGAVGSDI 60  
 QY 61 HPPDTPNFKGADSRVLAHRYGIVYKGYKLVNADVTIIAQAQKMLPHVGMANIAAD 120  
 DB 61 HPPDTPNFKGADSRVLAHRYGIVYKGYKLVNADVTIIAQAQKMLPHVGMANIAAD 120  
 QY 121 LETVDVFINVAKATTTKLGFEGRKEGIAVOAVLIER 157  
 DB 121 LEADVSQVAVKATITKLGFTGRABGIAQAVLILOK 157  
 RESULT 11  
 ISPF HABIN STANDARD; PRT; 158 AA.  
 AC P44815;  
 DT 01-NOV-1995 (Rel. 32, Created)  
 DT 01-NOV-1995 (Rel. 32, Last sequence update)

DT 15-SEP-2003 (Rel. 42, Last annotation update)  
 DB 2-C-methyl-D-erythritol 2,4-cyclodiphosphate synthase (EC 4.6.1.12)  
 DE (MECS) (MECDP-synthase).  
 GN ISPF OR H10671.  
 OS Haemophilus influenzae.  
 OC Pasteurellales; Pasteurellaceae; Haemophilus.  
 OC K0B1\_TaxId=727;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=RD / KW20 / ATCC 51907;  
 RX MEDLINE=95350630; PubMed=7542800;  
 RA Fleischmann R.D., Adams M.D., White O., Clayton R.A., Kirkness E.F.,  
 RA Kerlavage A.R., Bult C.J., Tomb J.-F., Dougherty B.A., Merrick J.M.,  
 RA McKenney K., Sutton G., Fitzhugh W., Fields C.A., Gocayne J.D.,  
 RA Scott J.D., Shirley R., Liu L.I., Glodek A., Kelley J.M.,  
 RA Weidman J.P., Phillips C.A., Spriggs T., Hedblom E., Cotton M.D.,  
 RA Ullendack T.R., Hanna M.C., Nguyen D.T., Sauke D.M., Brandon R.C.,  
 RA Fine L.D., Fritchman J.L., Geoghagen N.S.M.,  
 RA Gnehm C.L., McDonald L.A., Small K.V., Fraser C.M., Smith H.O.,  
 RA Venter J.C.;  
 RA "Whole-genome random sequencing and assembly of *Haemophilus influenzae*  
 RA *rd*,"  
 RT Science 269:496-512(1995).  
 RN [2]  
 RP IDENTIFICATION BY MASS SPECTROMETRY.  
 RX MEDLINE=98384070; PubMed=9719565;  
 RA Fountoulakis M., Duranville J.F., Roder D., Evers S., Berndt P.,  
 RA Langen H.;  
 RT "Reference map of the low molecular mass proteins of *Haemophilus*  
 RT *influenzae*,"  
 RT Electrophoresis 19:1819-1827(1998).  
 CC -1- FUNCTION: CONVERTS 4-DIPHOSPHOCYTIDYL-2C-METHYL-D-ERYTHRITOL 2-  
 CC PHOSPHATE INTO 2C-METHYL-D-ERYTHRITOL 2,4-CYCLODIPHOSPHATE AND  
 CC CMP. ALSO CONVERTS 4-DIPHOSPHOCYTIDYL-2C-METHYL-D-ERYTHRITOL INTO  
 CC 2C-METHYL-D-ERYTHRITOL 3,4-CYCLOPHOSPHATE AND CMP (BY SIMILARITY).  
 CC -1- CATALYTIC ACTIVITY: 2-phospho-4-(cytidine 5'-diphospho)-2-C-  
 CC methyl-D-erythritol = 2-C-methyl-D-erythritol 2,4-cyclodiphosphate  
 CC + CMP.  
 CC -1- PATHWAY: Nonaevonate terpenoid biosynthesis pathway; fifth step.  
 CC -1- SIMILARITY: BELONGS TO THE ISPF FAMILY.  
 CC -----  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation-  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
 CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
 CC -----  
 CC EMBL, U32750; AAC22331.1; --  
 CC PIR, F64156; F64156.  
 CC PDB; 1UN1; 21-AUG-02.  
 CC TIGR; H10671; --  
 CC HAMAP; MF 00107; -; 1.  
 CC InterPro; IPR003526; Ygbb.  
 CC Pfam; PF02542; Ygbb; 1.  
 CC TIGRFAWS; TIGR00151; ispf; 1.  
 CC PROSITE; PS01350; ispf; 1.  
 CC Lyase; Isoprene biosynthesis; Complete proteome; 3D-structure.  
 CC SEQUENCE 158 AA; 17194 MW; DC34BR347DEC2BF2 CRC64;  
 SQ  
 Query Match 66.3%; Score 534; DB 1; Length 158;  
 Best Local Similarity 65.6%; Pred. No. 5, 6e-43;  
 Matches 103; Conservative 20; Mismatches 34; Indels 0; Gaps 0;  
 QY 1 MIRVGMYDVARENDGHIILIGVKIPYKGLAHSDDGVVLAALADALIGAAALGDIK 60  
 DB 1 MERIGGFVHGVHVRPLILIGIEIPIYKGLGHSDAVVLAHTVADALGAVGSDI 60  
 QY 61 HPPDTPNFKGADSRVLAHRYGIVYKGYKLVNADVTIIAQAQKMLPHVGMANIAAD 120  
 DB 61 HPPDTPNFKGADSRVLAHRYGIVYKGYKLVNADVTIIAQAQKMLPHVGMANIAAD 120  
 QY 121 LETVDVFINVAKATTTKLGFEGRKEGIAVOAVLIER 157  
 DB 121 LEADVSQVAVKATITKLGFTGRABGIAQAVLILOK 157

QY 121 LETDNDPINKATTTETLGEGRKEGIAVOAVLIER 157  
 DB 121 LQCDIEQVNVKATTTETLGEGRKEGIAVOAVLIER 157

## RESULT 12

ISPF\_PASMTU STANDARD; PRT; 158 AA.

AC 157954;  
 DT 16-OCT-2001 (Rel. 40, Created)  
 DT 16-OCT-2001 (Rel. 40, Last sequence update)  
 DT 28-FEB-2003 (Rel. 41, Last annotation update)  
 DE 2-C-methyl-D-erythritol 2,4-cyclodiphosphate synthase (EC 4.6.1.12)  
 DE (MECPS) (MECDP-synthase).  
 GN ISPF OR PM1609.  
 OS Pasteurella multocida.  
 OC Bacteria; Proteobacteria; Gammaproteobacteria; Pasteurellales;  
 OC Pasteurellaceae; Pasteurella.  
 OX NCBI\_TaxID=747;  
 RN [1]

RP SEQUENCE FROM N.A.  
 RC STRAIN=PM70;  
 RA MEDLINE=2145866; PubMed=11248100;  
 RA May B., Zhang Q., Li L., Paustian M.L., Whitam T.S., Kapur V.;  
 RT "Complete genomic sequence of Pasteurella multocida PM70.";  
 RT Proc. Natl. Acad. Sci. U.S.A. 98:3460-3465(2001).  
 RL -1- FUNCTION: CONVERTS 4-DIPHOSPHOCYTIDYL-2C-METHYL-D-ERYTHRITOL 2-  
 PHOSPHATE INTO 2C-METHYL-D-ERYTHRITOL 2,4-CYCLODIPHOSPHATE AND  
 CMP. ALSO CONVERTS 4-DIPHOSPHOCYTIDYL-2C-METHYL-D-ERYTHRITOL INTO  
 2C-METHYL-D-ERYTHRITOL 3,4-CYCLOPHOSPHATE AND CMP (BY SIMILARITY).  
 CC -1- CATALYTIC ACTIVITY: 2-phospho-4-(cytidine 5'-diphospho)-2-C-  
 methyl-D-erythritol = 2-C-methyl-D-erythritol 2,4-cyclodiphosphate  
 + CMP.  
 CC -1- PATHWAY: Nonaevulonate terpenoid biosynthesis pathway; fifth step.  
 CC -1- SIMILARITY: BELONGS TO THE ISPF FAMILY.

CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
 CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).

CC EMBL: AB006198; AA03693.1; -  
 DR HAMAP: MF\_00107; -; 1.  
 DR InterPro: IPR003526; YqdB.  
 DR Pfam: PF02542; YqdB; 1.  
 DR TIGRFAMs: TIGR00151; ISPF; 1.  
 DR PROSITE: PS01350; ISPF; 1.  
 KW Lyase; Isoprene biosynthesis; Complete proteome.  
 SQ SEQUENCE 158 AA; 16555 MW; 35B9CBA59147171E CRC64;

Query Match 65.6%; Score 529; DB 1; Length 158;  
 Best Local Similarity 67.7%; Pred. No. 1.7e-42;  
 Matches 105; Conservative 15; Mismatches 35; Indels 0; Gaps 0;

QY 1 MIRVGQYDVHRENDGHIILGGVKKIPYKGLBAHSDGVDVLLHALADALIGAAALGDIGK 60  
 DB 1 MIRVGQYDVHRENDGHIILGGVKKIPYKGLBAHSDGVDVLLHALADALIGAAALGDIGK 60  
 QY 61 HPPDTPDNFKGADSRVLLRHVYGVYKKGKYLNVADVTIIAQAQKLPHVPGKRNIAAD 120  
 DB 61 HPPDTPDNFKGADSRVLLRHVYGVYKKGKYLNVADVTIIAQAQKLPHVPGKRNIAAD 120  
 QY 121 LETDNDPINKATTTETLGEGRKEGIAVOAVLIER 155  
 DB 121 LQCDIEQVNVKATTTETLGEGRKEGIAVOAVLIER 155

## RESULT 13

ISPF\_BACHD

ID ISPF\_BACHD STANDARD; PRT; 157 AA.

AC 09KGE7;  
 DT 16-OCT-2001 (Rel. 40, Created)  
 DT 16-OCT-2001 (Rel. 40, Last sequence update)  
 DT 28-FEB-2003 (Rel. 41, Last annotation update)  
 DE 2-C-methyl-D-erythritol 2,4-cyclodiphosphate synthase (EC 4.6.1.12)  
 DE (MECPS) (MECDP-synthase).  
 GN ISPF OR BE0108.  
 OS Bacillus halodurans.  
 OC Bacteria; Firmicutes; Bacillales; Bacillaceae; Bacillus.  
 OX NCBI\_TaxID=86665;  
 RN [1]

RP SEQUENCE FROM N.A.  
 RC STRAIN=C-125 / JCM 9153;  
 RA MEDLINE=20512582; PubMed=11058132;  
 RA Takami H., Nakasone K., Takaki Y., Mieno G., Sasaki R., Masui N.,  
 RA Fuji F., Hirama C., Nakamura Y., Ogasawara N., Kuhara S.,  
 RA Horikoshi K.;  
 RT "Complete genome sequence of the alkaliphilic bacterium Bacillus  
 RT halodurans and genomic sequence comparison with Bacillus subtilis.";  
 RT Nucleic Acids Res. 28:4317-4331(2000).  
 RL -1- FUNCTION: CONVERTS 4-DIPHOSPHOCYTIDYL-2C-METHYL-D-ERYTHRITOL 2-  
 PHOSPHATE INTO 2C-METHYL-D-ERYTHRITOL 2,4-CYCLODIPHOSPHATE AND  
 CMP. ALSO CONVERTS 4-DIPHOSPHOCYTIDYL-2C-METHYL-D-ERYTHRITOL INTO  
 2C-METHYL-D-ERYTHRITOL 3,4-CYCLOPHOSPHATE AND CMP (BY SIMILARITY).  
 CC -1- CATALYTIC ACTIVITY: 2-phospho-4-(cytidine 5'-diphospho)-2-C-  
 methyl-D-erythritol = 2-C-methyl-D-erythritol 2,4-cyclodiphosphate  
 + CMP.  
 CC -1- PATHWAY: Nonaevulonate terpenoid biosynthesis pathway; fifth step.  
 CC -1- SIMILARITY: BELONGS TO THE ISPF FAMILY.

CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
 CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).

CC EMBL: AP001507; BAB03827.1; -  
 DR PIR: D83663; D83663.  
 DR HAMAP: MF\_00107; -; 1.  
 DR InterPro: IPR003526; YqdB.  
 DR Pfam: PF02542; YqdB; 1.  
 DR TIGRFAMs: TIGR00151; ISPF; 1.  
 DR PROSITE: PS01350; ISPF; 1.  
 KW Lyase; Isoprene biosynthesis; Complete proteome.  
 SQ SEQUENCE 157 AA; 16751 MW; 0205862C7B65BA99 CRC64;

Query Match 64.5%; Score 520; DB 1; Length 157;  
 Best Local Similarity 66.5%; Pred. No. 1.1e-41;  
 Matches 103; Conservative 16; Mismatches 36; Indels 0; Gaps 0;

QY 1 MIRVGQYDVHRENDGHIILGGVKKIPYKGLBAHSDGVDVLLHALADALIGAAALGDIGK 60  
 DB 1 MIRVGQYDVHRENDGHIILGGVKKIPYKGLBAHSDGVDVLLHALADALIGAAALGDIGK 60  
 QY 61 HPPDTPDNFKGADSRVLLRHVYGVYKKGKYLNVADVTIIAQAQKLPHVPGKRNIAAD 120  
 DB 61 HPPDTPDNFKGADSRVLLRHVYGVYKKGKYLNVADVTIIAQAQKLPHVPGKRNIAAD 120  
 QY 121 LETDNDPINKATTTETLGEGRKEGIAVOAVLIER 155  
 DB 121 LQCDIEQVNVKATTTETLGEGRKEGIAVOAVLIER 155

## RESULT 14

ISPF\_NEIMB STANDARD; PRT; 160 AA.

AC 09JYM5;  
 DT 16-OCT-2001 (Rel. 40, Created)  
 DT 16-OCT-2001 (Rel. 40, Last sequence update)  
 DT 28-FEB-2003 (Rel. 41, Last annotation update)

DE 2-C-methyl-D-erythritol 2,4-cyclodiphosphate synthase (EC 4.6.1.12)  
 DE (MECPB) (MECDP-synthase).  
 GN ISPF OR NM31512.  
 OS Neisseria meningitidis (serogroup B).  
 OC Bacteria; Proteobacteria; Betaproteobacteria; Neisseriales;  
 OC Neisseriaceae; Neisseria.  
 NCBI\_TaxID=491;  
 RX SEQUENCE FROM N.A.  
 RC STRAIN=MC58 / Serogroup B;  
 MEDLINE=20175755; PubMed=10710307;  
 RA Tettelin H., Saunders N.C., Heidelberg J., Jeffries A.C., Nelson K.B.,  
 Eisen J.A., Ketchum K.A., Hood D.W., Heden J.F., Dodson R.J.,  
 Nelson W.C., Gwinn M.L., DeBoy R., Peterson J.D., Hickey E.K.,  
 Haft D.H., Salzberg S.L., White O., Fleischmann R.D., Dougherty B.A.,  
 Mason T., Ciecko A., Parksey D.S., Alair B., Clifton H., Clark E.B.,  
 Ra G.I.I., Scarlato V., Maignani V., Qin H., Vamathevan J.,  
 Salath H.O., Fraser C.M., Moxon E.R., Rappaport R., Venter J.C.,  
 "Complete genome sequence of Neisseria meningitidis serogroup B strain  
 MC58.";  
 RL Science 287:1809-1815(2000).  
 CC -1- FUNCTION: CONVERTS 4-DIPHOSPHOCYTIDYL-2C-METHYL-D-ERYTHRITOL 2-  
 PHOSPHATE INTO 2C-METHYL-D-ERYTHRITOL 2,4-CYCLODIPHOSPHATE AND  
 CMP. ALSO CONVERTS 4-DIPHOSPHOCYTIDYL-2C-METHYL-D-ERYTHRITOL INTO  
 2C-METHYL-D-ERYTHRITOL 3,4-CYCLOPHOSPHATE AND CMP (BY SIMILARITY).  
 CC -1- CATALYTIC ACTIVITY: 2-phospho-4-(cytidine 5'-diphospho)-2-C-  
 methyl-D-erythritol = 2-C-methyl-D-erythritol 2,4-cyclodiphosphate  
 + CMP.  
 CC -1- PATHWAY: Nonaevionate terpenoid biosynthesis pathway; fifth step.  
 CC -1- SIMILARITY: BELONGS TO THE ISPF FAMILY.  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 between the Swiss Institute of Bioinformatics and the EMBL outstation  
 at the European Bioinformatics Institute. There are no restrictions on its  
 use by non-profit institutions as long as its content is in no way  
 modified and this statement is not removed. Usage by and for commercial  
 entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
 or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
 CC EMBL; AE002501; AAF41868.1; -  
 DR PIR; G81073; G81073.  
 DR TIGR; NMB1512; -  
 DR HAMAP; MF\_00107; -; 1.  
 DR InterPro; IPR003526; Ygdb.  
 DR Pfam; PF02342; Ygdb; 1.  
 DR TIGRFAMs; TIGR00151; ISPF; 1.  
 DR PROSITE; PS01350; ISPF; 1.  
 KW Lyase; Isoprene biosynthesis; Complete proteome.  
 SQ SEQUENCE 160 AA; 17019 MW; 9C223BCED4D9BF3D CRC64;  
 Query Match 62.7%; Score 505; DB 1; Length 160;  
 Best Local Similarity 60.9%; Pred. No. 2.9e-40;  
 Matches 95; Conservative 21; Mismatches 40; Indels 0; Gaps 0;  
 QY 2 IRVWGVDVRFNDGHIILGGVPIPEKGLIAHSDDVYTHALADLIGAAAGDGGK 61  
 DB 4 IRVGGVDVRFNDGHIILGGVPIPEKGLIAHSDDVYTHALADLIGAAAGDGGK 63  
 QY 62 FPDTPNFKGADSRVLLRHVYGIYKGYKLVNADVTIIIAQAQKLPVPGMRANIADL 121  
 DB 64 FPDTPNFKGADSRVLLRHVYGIYKGYKLVNADVTIIIAQAQKLPVPGMRANIADL 123  
 QY 122 ETVDVFINVATTEKLGFBGREGIAVQAVYTLIER 157  
 DB 124 GIDISCVNIGKTKNEKDGFGREGIEAQAVALIVR 159

DT 30-MAY-2000 (Rel. 39, Last sequence update)  
 DT 28-FEB-2003 (Rel. 41, Last annotation update)  
 DE 2-C-methyl-D-erythritol 2,4-cyclodiphosphate synthase (EC 4.6.1.12)  
 DE (MECPB) (MECDP-synthase).  
 GN ISPF.  
 OS Haemophilus ducreyi.  
 OC Bacteria; Proteobacteria; Gammaproteobacteria; Pasteurellales;  
 OC Pasteurellaceae; Haemophilus.  
 NCBI\_TaxID=730;  
 RX SEQUENCE FROM N.A.  
 RC STRAIN=93000;  
 MEDLINE=96422469; PubMed=8825086;  
 RA Palmer K.P., Munson R.S. Jr.,  
 "Cloning and characterization of the genes encoding the hemolysin of  
 Haemophilus ducreyi.";  
 RL Mol. Microbiol. 18:871-830(1995).  
 CC -1- FUNCTION: CONVERTS 4-DIPHOSPHOCYTIDYL-2C-METHYL-D-ERYTHRITOL 2-  
 PHOSPHATE INTO 2C-METHYL-D-ERYTHRITOL 2,4-CYCLODIPHOSPHATE AND  
 CMP. ALSO CONVERTS 4-DIPHOSPHOCYTIDYL-2C-METHYL-D-ERYTHRITOL INTO  
 2C-METHYL-D-ERYTHRITOL 3,4-CYCLOPHOSPHATE AND CMP (BY SIMILARITY).  
 CC -1- CATALYTIC ACTIVITY: 2-phospho-4-(cytidine 5'-diphospho)-2-C-  
 methyl-D-erythritol = 2-C-methyl-D-erythritol 2,4-cyclodiphosphate  
 + CMP.  
 CC -1- PATHWAY: Nonaevionate terpenoid biosynthesis pathway; fifth step.  
 CC -1- SIMILARITY: BELONGS TO THE ISPF FAMILY.  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 between the Swiss Institute of Bioinformatics and the EMBL outstation  
 at the European Bioinformatics Institute. There are no restrictions on its  
 use by non-profit institutions as long as its content is in no way  
 modified and this statement is not removed. Usage by and for commercial  
 entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
 or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
 CC EMBL; U32175; AAC33539.1; -  
 DR PIR; S70845; S70845.  
 DR HAMAP; MF\_00107; -; 1.  
 DR InterPro; IPR003526; Ygdb.  
 DR Pfam; PF02542; Ygdb; 1.  
 DR TIGRFAMs; TIGR00151; ISPF; 1.  
 DR PROSITE; PS01350; ISPF; 1.  
 KW Lyase; Isoprene biosynthesis.  
 SQ SEQUENCE 160 AA; 17437 MW; D7F1F2F518F7000E CRC64;  
 Query Match 62.4%; Score 503; DB 1; Length 160;  
 Best Local Similarity 63.1%; Pred. No. 4.5e-40;  
 Matches 99; Conservative 17; Mismatches 41; Indels 0; Gaps 0;  
 QY 1 MTRVWGVDVRFNDGHIILGGVPIPEKGLIAHSDDVYTHALADLIGAAAGDGGK 60  
 DB 1 MTRVGGVDVRFNDGHIILGGVPIPEKGLIAHSDDVYTHALADLIGAAAGDGGK 60  
 QY 61 HPPDTPNFKGADSRVLLRHVYGIYKGYKLVNADVTIIIAQAQKLPVPGMRANIADL 120  
 DB 61 LPPDTPNFKGADSRVLLRHVYGIYKGYKLVNADVTIIIAQAQKLPVPGMRANIADL 120  
 QY 121 LETVDVFINVATTEKLGFBGREGIAVQAVYTLIER 157  
 DB 121 LNCDDVQINIVATTEKLGFBGREGIAVQAVYTLIER 157

Search completed: January 29, 2004, 15:50:41  
 Job time : 3.95645 secs

RESULT 15  
 ISPF HAEDU  
 ID ISPF HAEDU STANDARD; PRT; 160 AA.  
 AC 047956;  
 DT 30-MAY-2000 (Rel. 39, Created)

GenCore version 5.1.6  
Copyright (c) 1993 - 2004 Compugen Ltd.

OM protein - protein search, using sw model

Run on: January 29, 2004, 15:44:14 ; Search time 11.7026 Seconds  
(without alignments)  
3461.979 Million cell updates/sec

Title: US-09-941-947A-14

Perfect score: 806  
Sequence: 1 MIRVGMGYDVHRENDGDHIL.....LAFEGRGKGIWAQAVLIER 157

Scoring table: BLOSUM62  
Gapop 10.0 , Gapext 0.5

Searched: 830525 segs, 258052604 residues

Total number of hits satisfying chosen parameters: 830525

Minimum DB seq length: 0  
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Database :

SPTREMBL.23.\*  
1: sp\_archaea.\*  
2: sp\_bacteria.\*  
3: sp\_fungi.\*  
4: sp\_human.\*  
5: sp\_invertebrate.\*  
6: sp\_mammal.\*  
7: sp\_mhc.\*  
8: sp\_organelle.\*  
9: sp\_phase.\*  
10: sp\_plant.\*  
11: sp\_rodent.\*  
12: sp\_virus.\*  
13: sp\_vertebrate.\*  
14: sp\_unclassified.\*  
15: sp\_virus.\*  
16: sp\_bacteriophage.\*  
17: sp\_archaeop.\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

# SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	569	70.6	159	16	Q8EBR3 Q8EBR3 shewanella
2	551	68.4	158	16	Q8DCS9 Q8DCS9 vibrio vuln
3	537	66.6	159	16	Q8FEG6 Q8FEG6 escherichia
4	453	56.2	157	16	Q8KCS5 Q8KCS5 chlorobium
5	406	50.4	151	16	Q8GDM4 Q8GDM4 bruceella su
6	394	48.9	161	16	Q8DHG4 Q8DHG4 synchococc
7	363	45.0	161	16	Q8FDM5 Q8FDM5 leptospira
8	329.5	40.9	157	16	Q8D224 Q8D224 wigglesworth
9	272	33.7	152	2	Q8KQF5 Q8KQF5 thermus the
10	229	28.4	174	16	Q8G5J2 Q8G5J2 bifidobacte
11	208.5	25.9	171	16	Q8FMI4 Q8FMI4 corynebacte
12	194	24.1	165	16	Q8GUA4 Q8GUA4 mycoplasma
13	180	22.3	89	2	Q8KXO6 Q8KXO6 synchococc
14	87.5	10.9	571	16	Q8EPR7 Q8EPR7 shewanella
15	86	10.7	434	10	Q8LBY9 Q8LBY9 arabidopsis
16	85	10.5	377	16	Q8P721 Q8P721 xanthomonas

17	85	10.5	434	10	Q93X74 Q93X74 brassica ra
18	84.5	10.5	429	16	Q8EBJ7 Q8EBJ7 shewanella
19	84.5	10.5	748	16	Q9KZ42 Q9KZ42 streptomyc
20	84	10.4	748	16	Q8XUE8 Q8XUE8 ralsomona s
21	83.5	10.4	410	16	Q8B722 Q8B722 streptococc
22	83.5	10.4	410	16	Q8B722 Q8B722 streptococc
23	83	10.3	373	16	Q8B722 Q8B722 streptococc
24	83	10.3	593	5	Q8PJE0 Q8PJE0 xanthomonas
25	82.5	10.2	767	16	Q92T08 Q92T08 xanthomonas
26	82	10.2	939	16	Q9WXT0 Q9WXT0 xanthomonas
27	82	10.2	943	16	Q8NQ98 Q8NQ98 corynebacte
28	81.5	10.1	727	16	Q9KN35 Q9KN35 vibrio chol
29	81.5	10.1	727	16	Q9KEV1 Q9KEV1 vibrio chol
30	81	10.0	434	10	Q9LPA3 Q9LPA3 arabidopsis
31	80.5	10.0	203	16	Q8PAI1 Q8PAI1 escherichia
32	80	9.9	209	16	Q8XK15 Q8XK15 clostridium
33	80	9.9	319	17	Q970V2 Q970V2 sulfobobus
34	80	9.9	383	16	Q8UE77 Q8UE77 agrobacteri
35	80	9.9	444	10	Q8GVP6 Q8GVP6 oryza sativ
36	79.5	9.9	452	17	Q970M5 Q970M5 sulfobobus
37	79.5	9.9	941	16	Q8FPA8 Q8FPA8 corynebacte
38	79.5	9.9	5206	16	Q8DEP9 Q8DEP9 vibrio vuln
39	79	9.8	903	12	Q64928 Q64928 african hor
40	78.5	9.7	513	16	Q988N0 Q988N0 rhizobium 1
41	78.5	9.7	734	16	Q988N0 Q988N0 rhizobium 1
42	78	9.7	431	13	Q05467 Q05467 gallus gall
43	78	9.7	718	16	Q97EP8 Q97EP8 clostridium
44	78	9.7	1840	13	Q90831 Q90831 gallus gall
45	77.5	9.6	173	16	Q8FPR6 Q8FPR6 corynebacte

## ALIGNMENTS

### RESULT 1

ID	Q8EBR3	PRELIMINARY:	PRT:	159 AA.
AC	Q8EBR3	01-MAR-2003 (TREMBL:rel. 23, Created)		
DT	01-MAR-2003 (TREMBL:rel. 23, Last sequence update)			
DT	01-MAR-2003 (TREMBL:rel. 23, Last annotation update)			
DE	2C-methyl-D-erythritol 2,4-cyclodiphosphate synthase.			
DN	1SPF OR S03437			
OS	Shewanella oneidensis.			
OC	Bacteria; Proteobacteria; Gammaproteobacteria; Alteromonadales;			
OC	Alteromonadaceae; Shewanella.			
OX	NCBI_TaxID=70863;			
RN	[1]			
RP	SEQUENCE FROM N.A.			
RC	STRAIN=MR-1;			
RX	MEDLINE=22297686; PubMed=12368813;			
RA	Heidelberg J.F., Paulsen I.T., Nelson K.E., Gaidos E.J., Nelson W.C.,			
RA	Read T.D., Eisen J.A., Seshadri R., Ward N., Meehe B., Clayton R.A.,			
RA	Meyer T., Tsaplin A., Scott J., Beanan M., Brinkac L., Daugherty S.,			
RA	DeBoy R.T., Dodson R.J., Durkin A.S., Haft D.H., Kolonay J.F.,			
RA	Madupu R., Peterson J.D., Umayam L.A., White O., Wolf A.M.,			
RA	Varatharajan J., Weidman J., Impraim M., Lee K., Berry K., Lee C.,			
RA	Mueller J., Khouri H., Gill J., Uitterback T.R., McDonald L.A.,			
RA	Feldblyum T.V., Smith H.O., Venter J.C., Neilson K.H., Frazer C.M.,			
RT	"Genome sequence of the dissimilatory metal ion-reducing bacterium			
RT	Shewanella oneidensis."			
RL	Nat. Biotechnol. 20:1118-1123(2002).			
DR	EMBL; AB015780; AAM56434.1; -			
DR	TIGR; S03437; -			
KW	Complete proteome.			
SQ	SEQUENCE 159 AA; 16995 MW; A16DC82586297501 CRC64;			

Query Match 70.6%; Score 569; DB 16; Length 159;

Best Local Similarity 71.2%; Pred. No. 2,7e-45;

Matches 111; Conservative 17; Mismatches 28; Indels 0; Gaps 0;

2 IRVGMGYDVHRENDGDHILIGSVKIPYKGLHAGSDGVVTLALADILALAAAGTGK 61

DB 3 IRIGGFVHKGGEPRPILICGVYEVYKELVANSDDVVLHAIADALIGAGADIGKH 62  
 QY 62 FPDTPNFKGADSRVLLRHYVGIYKKGKYLVAADVTIIAQAAPKMLPHVPGKRAMIAADL 121  
 DB 63 FPDTPNFKGADSRVLLRHYVGIYKKGKYLVAADVTIIAQAAPKMLPHVPGKRAMIAADL 122  
 QY 122 ETDVDFINVKATTEKLGFGKRGKGIYAQAVALIER 157  
 DB 123 NADVADINVKATTEKLGFGKRGKGIYAQAVALIER 158

## RESULT 2

Q8BC59 PRELIMINARY; PRT; 158 AA.  
 AC Q8BC59;  
 DT 01-MAR-2003 (TREMELREL. 23, Created)  
 DT 01-MAR-2003 (TREMELREL. 23, Last sequence update)  
 DT 01-MAR-2003 (TREMELREL. 23, Last annotation update)  
 DE 2C-methyl-D-erythritol 2,4-cyclodiphosphate synthase.  
 GN VLI1583.  
 OS Vibrrio vulnificus.  
 OC Bacteria; Proteobacteria; Gammaproteobacteria; Vibrionales;  
 OC Vibrionaceae; Vibrrio.  
 OX NCBI\_Taxid=672;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=CNC6;  
 RA Rhee J.H., Kim S.Y., Chung S.S., Kim J.J., Moon Y.H., Jeong H.,  
 RA Choy H.E.;  
 RT "Complete genome sequence of *Vibrio vulnificus* CNC6."  
 RL Submitted (DEC-2002) to the EMBL/GenBank/DBJ databases.  
 DR EMBL; AE016802; AMO1006.1; -  
 KW Complete proteome.  
 SQ SEQUENCE 158 AA; 17094 MW; 0B888FE050160112 CRC64;

Query Match 68.4%; Score 551; DB 16; Length 158;  
 Best Local Similarity 69.0%; Pred. No. 1.3e-43;  
 Matches 107; Conservative 19; Mismatches 29; Indels 0; Gaps 0;

QY 1 MIRVGKGVHRRFNDGDHIIIGGVKIPYKGLAHSDDGVVLHAIADALIGAGADIGKH 60  
 DB 1 MRIGGFVHKGGEPRPILICGVYEVYKELVANSDDVVLHAIADALIGAGADIGKH 60  
 QY 61 FPDTPNFKGADSRVLLRHYVGIYKKGKYLVAADVTIIAQAAPKMLPHVPGKRAMIAAD 120  
 DB 61 FPDTPNFKGADSRVLLRHYVGIYKKGKYLVAADVTIIAQAAPKMLPHVPGKRAMIAAD 120  
 QY 121 LETVDVFINVKATTEKLGFGKRGKGIYAQAVALIER 155  
 DB 121 LETVDVFINVKATTEKLGFGKRGKGIYAQAVALIER 155

## RESULT 3

Q8BC59 PRELIMINARY; PRT; 159 AA.

AC Q8BC59;  
 DT 01-MAR-2003 (TREMELREL. 23, Created)  
 DT 01-MAR-2003 (TREMELREL. 23, Last sequence update)  
 DT 01-MAR-2003 (TREMELREL. 23, Last annotation update)  
 DE 2-C-methyl-D-erythritol 2,4-cyclodiphosphate synthase (BC  
 4.6.1.12).  
 GN YGBB OR C3913.  
 OS Escherichia coli O6.  
 OC Bacteria; Proteobacteria; Gammaproteobacteria; Enterobacteriales;  
 OC Enterobacteriaceae; Escherichia.  
 OX NCBI\_Taxid=217992;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=C6:H1 / CFT073 / ATCC 700928;  
 RA MEDLINE=22388234; PubMed=12471157;  
 RA Welch R.A., Burland V., Plunkett G. III, Redford P., Soesch P.,  
 RA Rasko D., Buckles E.L., Liu S.-R., Boutin A., Hackett J., Stroud D.,  
 RA Mayhew G.F., Rose D.J., Zhou S., Schwartz D.C., Perna N.T.;

RA Mobley H.L.T., Donenberg M.S., Blattner F.R.;  
 RT "Extensive mosaic structure revealed by the complete genome sequence  
 of uropathogenic *Escherichia coli*."  
 RL Proc. Natl. Acad. Sci. U.S.A. 99:17020-17024 (2002).  
 DR EMBL; AB016765; AAB01762.1; -  
 KW Lyase; Complete proteome.  
 SQ SEQUENCE 159 AA; 16867 MW; 8FD4573623BA4549 CRC64;

Query Match 66.6%; Score 537; DB 16; Length 159;  
 Best Local Similarity 69.5%; Pred. No. 2.6e-42;  
 Matches 107; Conservative 14; Mismatches 33; Indels 0; Gaps 0;

QY 2 IRVKGVDVHRRFNDGDHIIIGGVKIPYKGLAHSDDGVVLHAIADALIGAGADIGKH 61  
 DB 1 MRIGGFVHKGGEPRPILICGVYEVYKELVANSDDVVLHAIADALIGAGADIGKH 60  
 QY 62 FPDTPNFKGADSRVLLRHYVGIYKKGKYLVAADVTIIAQAAPKMLPHVPGKRAMIAAD 121  
 DB 61 FPDTPNFKGADSRVLLRHYVGIYKKGKYLVAADVTIIAQAAPKMLPHVPGKRAMIAAD 120  
 QY 122 ETDVDFINVKATTEKLGFGKRGKGIYAQAVALIER 155  
 DB 121 GCHMDVFINVKATTEKLGFGKRGKGIYAQAVALIER 154

## RESULT 4

Q8BC25 PRELIMINARY; PRT; 157 AA.

AC Q8BC25;  
 DT 01-OCT-2002 (TREMELREL. 22, Created)  
 DT 01-OCT-2002 (TREMELREL. 22, Last sequence update)  
 DT 01-MAR-2003 (TREMELREL. 23, Last annotation update)  
 DE 2C-methyl-D-erythritol 2,4-cyclodiphosphate synthase.  
 GN ISPP OR CT1601.  
 OS Chlorobium tepidum.  
 OC Bacteria; Chlorobi; Chlorobia; Chlorobiales; Chlorobiaceae;  
 OC Chlorobiaceae; Chlorobium.  
 OX NCBI\_Taxid=1097;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=TJS / ATCC 49652 / DSM 12025;  
 RA MEDLINE=22103685; PubMed=12093901;  
 RA Eisen J.A., Nelson K.R., Paulsen I.T., Heidelberg J.F., Wu M.,  
 RA Dodson R.J., Deboy R., Gwinn M.L., Nelson W.C., Haft D.H.,  
 RA Hickey R.K., Peterson J.D., Durkin A.S., Kolonay J.L., Yang F.,  
 RA Holt I., Umayam L.A., Mason T., Brenner M., Shea T.P., Parksey D.,  
 RA Niemeyer W.C., Feldblyum T.V., Hansen C.L., Craven W.B., Radu D.,  
 RA Vamathevan J., Khouli H., White O., Gruber T.M., Ketchum K.A.,  
 RA Venter J.C., Teitelin H., Bryant D.A., Fraser C.M.;  
 RT "The complete genome sequence of *Chlorobium tepidum* TJS, a  
 RT photosynthetic, anaerobic, green-sulfur bacterium."  
 RL Proc. Natl. Acad. Sci. U.S.A. 99:9509-9514 (2002).  
 DR EMBL; AB012915; AAM72826.1; -  
 DR TIGR; CT1601; -  
 DR InterPro; IPR003526; Ygbb.  
 DR Pfam; PF02542; Ygbb; 1.  
 DR TIGRPFAMs; TIGR00151; ISPP; 1.  
 DR PROSITE; PS01350; ISPP; 1.  
 KW Complete proteome.  
 SQ SEQUENCE 157 AA; 16923 MW; 4F5A7DCB20C897B CRC64;

Query Match 56.2%; Score 453; DB 16; Length 157;  
 Best Local Similarity 54.2%; Pred. No. 1.8e-38;  
 Matches 84; Conservative 33; Mismatches 38; Indels 0; Gaps 0;

QY 2 IRVKGVDVHRRFNDGDHIIIGGVKIPYKGLAHSDDGVVLHAIADALIGAGADIGKH 61  
 DB 1 MRIGGFVHKGGEPRPILICGVYEVYKELVANSDDVVLHAIADALIGAGADIGKH 60  
 QY 62 FPDTPNFKGADSRVLLRHYVGIYKKGKYLVAADVTIIAQAAPKMLPHVPGKRAMIAAD 121  
 DB 61 FPDTPNFKGADSRVLLRHYVGIYKKGKYLVAADVTIIAQAAPKMLPHVPGKRAMIAAD 120

QY 122 ETDVDFINVKATTEKLGFEGRKEGIAVQAVLIE 156  
 Db 121 GLEINAVSVATTTNEKLGFGVGRGACAAVVCIE 155

## RESULT 5

Q8DGH4 PRELIMINARY; PRT; 451 AA.

AC Q8DGH4; PRELIMINARY; PRT; 451 AA.  
 DT 01-MAR-2003 (TREMBlrel. 23, Created)  
 DT 01-MAR-2003 (TREMBlrel. 23, Last sequence update)  
 DE 2C-methyl-D-erythritol 2,4-cyclodiphosphate  
 DE synthase/4-diphosphocytidylyl-2C-methyl-D-erythritol synthase.  
 BR1120.  
 OS Brucella suis.  
 OC Bacteria; Proteobacteria; Alphaproteobacteria; Rhizobiales;  
 OC Brucellaceae; Brucella.  
 NCBI\_TaxID=29461;  
 RX MEDLINE=2224741; PubMed=12271122;  
 RA Paulsen I.T., Seshadri R., Nelson K.E., Eisen J.A., Heidelberg J.F.,  
 RA Read T.D., Dodson R.J., Umayam L., Brinkac L.M., Beanan M.J.,  
 RA Daugherty S.C., Deboy R.T., Durkin A.S., Kolonay J.F., Madupu R.,  
 RA Nelson W.C., Ayodeji B., Krahl M., Shetty J., Malek J., Van Aken S.E.,  
 RA Riedmuller S., Tettelin H., Gill S.R., White O., Salzberg S.L.,  
 RA Hoover D.L., Lindler L.B., Halling S.M., Boyle S.M., Fraser C.M.;  
 RT "The Brucella suis genome reveals fundamental similarities between  
 animal and plant pathogens and symbionts";  
 Proc. Natl. Acad. Sci. U.S.A. 99:13148-13153(2002).  
 DR EMBL, AF014413; AAN30040.1; -.  
 DR TIGR, BR1120; -.  
 SQ COMPLETE proteome.  
 SEQ 451 AA; 48492 MW; 3BDB7285C056A161 CRC64;

Query Match 50.4%; Score 406; DB 16; Length 451;  
 Best Local Similarity 51.3%; Pred. No. 1.7e-29;  
 Matches 79; Conservative 20; Mismatches 55; Indels 0; Gaps 0;

QY 2 IRVNGYDVHFRNDGHIILGKVKIPYKGLFAHSDGCVLHALADAILGAALGDIKH 61  
 Db 290 IRTGNGYDVHFRNDGHIILGKVKIPYKGLFAHSDGCVLHALADAILGAALGDIKH 349  
 QY 62 PPDTPDNFKGADSRVLAHNVGYIVKEKGYKLVNADVTIIAQAAPKPLPHVPGGRANIAADL 121  
 Db 350 PPDTPDNFKGADSRVLAHNVGYIVKEKGYKLVNADVTIIAQAAPKPLPHVPGGRANIAADL 409  
 QY 122 ETDVDFINVKATTEKLGFEGRKEGIAVQAVLIE 155  
 Db 410 GLEINAVSVATTTNEKLGFGVGRGACAAVVCIE 443

## RESULT 6

Q8DHC4 PRELIMINARY; PRT; 161 AA.

AC Q8DHC4; PRELIMINARY; PRT; 161 AA.  
 DT 01-MAR-2003 (TREMBlrel. 23, Created)  
 DT 01-MAR-2003 (TREMBlrel. 23, Last sequence update)  
 DE 2C-methyl-D-erythritol 2,4-cyclodiphosphate synthase.  
 TIR2035.  
 GN Synecchococcus elongatus (Thermosynechococcus elongatus).  
 OS Bacteria; Cyanobacteria; Chroococcales; Synecchococcus.  
 NCBI\_TaxID=32046;  
 RX MEDLINE=2225144; PubMed=12240834;  
 RA Nakamura Y., Kaneo T., Sato S., Ikeuchi M., Katoh H., Sasamoto S.,  
 RA Matsumoto M., Matsumoto M., Nakazaki N.,  
 RA Kiyokawa C., Kohara M., Matsumoto M., Nakazaki N.,

RA Shimo S., Sugimoto M., Takeuchi C., Yamada M., Tabata S.;  
 RT "Complete genome structure of the thermophilic cyanobacterium  
 RT Thermosynechococcus elongatus BP-1";  
 RL DNA Res. 9:123-130(2002).  
 DR EMBL, AF005376; BAC09587.1; -.  
 KW Complete proteome.  
 SQ SEQUENCE 161 AA; 17323 MW; CA4C67BB6D02498D CRC64;

Query Match 48.9%; Score 394; DB 16; Length 161;  
 Best Local Similarity 45.9%; Pred. No. 5.8e-29;  
 Matches 75; Conservative 33; Mismatches 47; Indels 0; Gaps 0;

QY 2 IRVNGYDVHFRNDGHIILGKVKIPYKGLFAHSDGCVLHALADAILGAALGDIKH 61  
 Db 4 IRTGNGYDVHFRNDGHIILGKVKIPYKGLFAHSDGCVLHALADAILGAALGDIKH 63  
 QY 62 PPDTPDNFKGADSRVLAHNVGYIVKEKGYKLVNADVTIIAQAAPKPLPHVPGGRANIAADL 121  
 Db 64 PPDTPDNFKGADSRVLAHNVGYIVKEKGYKLVNADVTIIAQAAPKPLPHVPGGRANIAADL 123  
 QY 122 ETDVDFINVKATTEKLGFEGRKEGIAVQAVLIE 156  
 Db 124 GLEINAVSVATTTNEKLGFGVGRGACAAVVCIE 158

## RESULT 7

Q8FOA5 PRELIMINARY; PRT; 161 AA.

AC Q8FOA5; PRELIMINARY; PRT; 161 AA.  
 DT 01-MAR-2003 (TREMBlrel. 23, Created)  
 DT 01-MAR-2003 (TREMBlrel. 23, Last sequence update)  
 DE YGDB family protein.  
 GN YGDB OR LA3591.  
 OS Leptospira interrogans.  
 OC Bacteria; Spirochaetes; Spirochaetales; Leptospiraceae; Leptospira.  
 NCBI\_TaxID=173;  
 RX MEDLINE=56601; Serogroup Icterohaemorrhagiae / Serovar 1a;  
 RA Ren S.;  
 RA Submitted (MAR-2002) to the EMBL/GenBank/DBJ databases.  
 DR EMBL, AF011514; AAN50789.1; -.  
 KW Complete proteome.  
 SQ SEQUENCE 161 AA; 17651 MW; 625A49B3223022C9 CRC64;

Query Match 45.0%; Score 363; DB 16; Length 161;  
 Best Local Similarity 45.9%; Pred. No. 4.5e-26;  
 Matches 73; Conservative 32; Mismatches 52; Indels 2; Gaps 1;

QY 1 MIRVNGYDVHFRNDGHIILGKVKIPYKGLFAHSDGCVLHALADAILGAALGDI 58  
 Db 1 MYRIGNGYDVHFRNDGHIILGKVKIPYKGLFAHSDGCVLHALADAILGAALGDI 60  
 QY 59 GKPPDTPDNFKGADSRVLAHNVGYIVKEKGYKLVNADVTIIAQAAPKPLPHVPGGRANIA 118  
 Db 61 GKPPDTPDNFKGADSRVLAHNVGYIVKEKGYKLVNADVTIIAQAAPKPLPHVPGGRANIA 120  
 QY 119 ADLETDVDFINVKATTEKLGFEGRKEGIAVQAVLIE 157  
 Db 121 MLNMLPDCISVYKATTEKLGFGVGRGACAAVVCIE 159

## RESULT 8

Q8D224 PRELIMINARY; PRT; 157 AA.

AC Q8D224; PRELIMINARY; PRT; 157 AA.  
 DT 01-MAR-2003 (TREMBlrel. 23, Created)  
 DT 01-MAR-2003 (TREMBlrel. 23, Last sequence update)  
 DE YGDB protein.  
 GN YGDB.  
 OS Wigglesworthia brevipalpis.

OC Bacteria; Proteobacteria; Gammaproteobacteria; Enterobacteriales;  
 OC Enterobacteriaceae; Wiglesworthia.  
 OC NCBI\_TaxID=164609;  
 RN (1)  
 RP SEQUENCE FROM N.A.  
 RX MEDLINE=22297718; PubMed=12219091;  
 RA Akman L., Yamashita A., Matanabe H., Oshima K., Shiba T., Hattori M.,  
 RA Aksoy S.;  
 RT "Genome sequence of the endocellular obligate symbiont of tsetse  
 RT flies, *Wiglesworthia glossinidia*."  
 RL Nat. Genet. 32:402-407 (2002).  
 DR EMBL; AB063522; BAC24677.1; -  
 KW Complete proteome.  
 SQ SEQUENCE 157 AA; 17252 MW; 775DCB1DDAE2884A CRC64;

Query Match 40.9%; Score 329.5; DB 16; Length 157;  
 Best Local Similarity 45.2%; Pred. No. 5.8e-23;

Matches 71; Conservative 37; Mismatches 46; Indels 3; Gaps 3;

QY 2 IRVGMGYDVHFRND-GDHIILGCVKIPYKGLBAHSDGVVLAHALADALIGAAALGDIGK 60  
 DB 1 MRVGHGDFHFKFKIKYKPLILGCVHPIPYCKGVVSHSDGVIIHSILSLGASSLGDICI 60  
 QY 61 HEPDTPNFGAASRVLRHYVGIYKKGKLVNADVTIIAQPXKLPHVPGKRAIIAAD 120  
 DB 61 LFPNNDIKYKINISCLLQVWNCIKKK-YKIGNIDVTLLFLEYPKISSTYKICICISNC 119  
 QY 121 LETDVPINVKATTTKLGFEGRKGIAGVAV- VLE 156  
 DB 120 LKCKTIVINIKSKTMGLGIGKKGLASAVSLLE 156

# RESULT 9 Q8R0PS

PRELIMINARY; PRT; 152 AA.

ID Q8R0PS  
 DT 01-JUN-2002 (T-EMBLrel. 21, Created)  
 DT 01-JUN-2002 (T-EMBLrel. 21, Last sequence update)  
 DT 01-OCT-2002 (T-EMBLrel. 22, Last annotation update)  
 DE 2-C-methyl-D-erythritol 2,4-cyclodiphosphate synthase  
 DE (EC 4.6.1.12).  
 GN YGBB.  
 OS Thermus thermophilus.  
 OC Bacteria; Deinococcus-Thermus; Deinococci; Thermales; Thermaceae;  
 OC Thermus.  
 OC NCBI\_TaxID=274;  
 RN (1)  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=HB8;  
 RA Kishida H., Wada T., Unzai S., Kuzuyama T., Terada T., Shirouzu M.,  
 RA Yokoyama S., Tame J.R., Park S.;  
 RT "2-C-methyl-D-erythritol 2,4-cyclodiphosphate synthase from *Thermus*  
 RT thermophilus HB8."  
 RL Submitted (MAR-2002) to the EMBL/GenBank/DBJ databases.  
 DR EMBL; AB082126; BAB86885.1; -  
 DR InterPro: IPR003526; YGBB.  
 DR Pfam: PF02542; YGBB; 1.  
 DR TIGRPFAM; TIGR00151; ISPF; 1.  
 DR PROSITE; PS01350; ISPF; 1.  
 KW Lyase.  
 SQ SEQUENCE 152 AA; 16520 MW; 0114C0B44DC28F3 CRC64;

Query Match 33.7%; Score 272; DB 2; Length 157;  
 Best Local Similarity 39.4%; Pred. No. 1.3e-17;

Matches 61; Conservative 25; Mismatches 65; Indels 4; Gaps 1;

QY 2 IRVGMGYDVHFRNDGHIILGCVKIPYKGLBAHSDGVVLAHALADALIGAAALGDIGK 61  
 DB 1 MRVGHGDFHFKFKIKYKPLILGCVHPIPYCKGVVSHSDGVIIHSILSLGASSLGDICI 60  
 QY 62 FPDTPNFGAASRVLRHYVGIYKKGKLVNADVTIIAQPXKLPHVPGKRAIIAADL 121  
 DB 61 LFPDTPNFGAASRVLRHYVGIYKKGKLVNADVTIIAQPXKLPHVPGKRAIIAADL 120

QY 122 ETDVDFINVKATTTKLGFEGRKGIAGVAVVLE 156  
 DB 121 RLPDRIIGLTFKTSGLA-----PSHVQARAVVLLD 151

## RESULT 10

PRELIMINARY; PRT; 174 AA.

ID Q8G5L2  
 DT 01-MAR-2003 (T-EMBLrel. 23, Created)  
 DT 01-MAR-2003 (T-EMBLrel. 23, Last sequence update)  
 DT 01-MAR-2003 (T-EMBLrel. 23, Last annotation update)  
 DE 2-C-methyl-D-erythritol 2,4-cyclodiphosphate synthase, Mecsps.  
 DE Mecsps-synthase.  
 GN ISPF OR BL0997.  
 OS Bifidobacterium longum.  
 OC Bacteria; Actinobacteriae; Actinobacteriales; Bifidobacteriales;  
 OC Bifidobacteriaceae; Bifidobacterium.  
 OC NCBI\_TaxID=216816;  
 RN (1)  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=NCC 2705;  
 RX MEDLINE=22294977; PubMed=12381787;  
 RA Scheil M.A., Karimantou M., Snel B., Vilanova D., Berger B.,  
 RA Pessi G., Zwahlen M.-C., Desiere F., Bork P., Delley M.,  
 RA Pridmore R.D., Arigoni F.;  
 RT "The genome sequence of *Bifidobacterium longum* reflects its adaptation  
 RT to the human gastrointestinal tract."  
 RL Proc. Natl. Acad. Sci. U.S.A. 99:14422-14427 (2002).  
 DR EMBL; AE014722; AAN24805.1; -  
 KW Complete proteome.  
 SQ SEQUENCE 174 AA; 17433 MW; F02406B3DBE336B CRC64;

Query Match 28.4%; Score 229; DB 16; Length 174;  
 Best Local Similarity 37.0%; Pred. No. 1.6e-13;

Matches 64; Conservative 20; Mismatches 67; Indels 22; Gaps 6;

QY 2 IRVGMGYDVHFRNDGHIILGCVKIP-----YKGLBAHSDGVVLAHALA 46  
 DB 6 VLIQGFPAHFRFAPAGSSRELIMINGLVWPDSCSBADMAAYE-GIEDSDGVVAAHALI 64  
 QY 47 DALIGAAALGDIGKRP-PTDTPNFGAASRVLRHYVGIYKKGKLVNADVTIIAQPX 105  
 DB 65 DALIAAARLGDIGSLFGVADAGAGMGHIDMLDEYVAHALASNGYTPASASVALIIGKRP 124  
 QY 106 MLPVPGKRAIIAADLETDVPF-INVKATTTKLGFEGRKGIAGVAVVLE 157  
 DB 125 -----IGTRBAABAALSAAGCPVSVTNTTDDHKGFTGKSGIAAIAALVLEK 173

## RESULT 11

PRELIMINARY; PRT; 171 AA.

ID Q8FM14  
 DT 01-MAR-2003 (T-EMBLrel. 23, Created)  
 DT 01-MAR-2003 (T-EMBLrel. 23, Last sequence update)  
 DT 01-MAR-2003 (T-EMBLrel. 23, Last annotation update)  
 DE Conserved hypothetical protein.  
 GN CE2520.  
 OS Corynebacterium efficiens.  
 OC Bacteria; Actinobacteriae; Actinobacteriales; Actinomycetales;  
 OC Corynebacteriaceae; Corynebacteriaceae; Corynebacterium.  
 OC NCBI\_TaxID=152794;  
 RN (1)  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=YS-314 / AJ 12310 / DSM 44549 / JCM 11189;  
 RA Kawarabayashi Y., Yamazaki J., Hino Y., Kikuchi H., Nakamura Y.,  
 RA Ikeo K., Suzuki M., Mashima J., Itoh T., Yamagishi A., Nishio Y.,  
 RA Ueda Y., Sugimoto S.;  
 RT "The entire genomic sequence of *Corynebacterium efficiens* YS-314."  
 RL Submitted (MAY-2002) to the EMBL/GenBank/DBJ databases.  
 DR EMBL; AP005222; BAC19330.1; -

KM Hypothetical protein; Complete proteome.  
SQ SEQUENCE 171 AA; 18076 MW; 524D47207A085EB7 CRC64;

Query Match 25.9%; Score 208.5; DB 16; Length 171;  
Best Local Similarity 33.1%; Pred. No. 1.2e-11;  
Matches 52; Conservative 25; Mismatches 69; Indels 11; Gaps 3;

QY 3 RVGMGYDVHRENDGHIILGVKIPYEKGLAHSDDGVVLAADALIGAAALDIGHK 62  
DB RVGIASDAHQIEAGKFCMICKLFFDVGDCGEGSDGVVAHAIYDALISASLGDSFV 78  
QY 63 PDDDPFKADSRVLLRHVYGVYKKGKYLVAADVTTIIAQAPKMLPHVGMKANIADIE 122  
DB GVGRPEYDVGSGTQLLEVAELLITRGFTIGVAAQLVQGPFR-----GPRREAOQV 133  
QY 123 TDV----DFINVAATTEKLGFEGRKEGIAVQAVLI 155  
DB SDILGAPCFHS--ATTDDHMGFTGRGEGRAALATAV 168

## RESULT 12

Q8EUA4 PRELIMINARY; PRT; 165 AA.

AC Q8EUA4; 01-MAR-2003 (TREMBlrel. 23, Created)  
DT 01-MAR-2003 (TREMBlrel. 23, Last sequence update)  
DE Putative enzyme of deoxy-xylulose pathway ygbB.  
GN MYPEI0270.  
OS Mycoplasma penetrans.  
OC Bacteria; Firmicutes; Mollicutes; Mycoplasmataceae; Mycoplasma.  
OX NCBI\_TaxID=28227;  
RN [1]  
RP SEQUENCE FROM N.A.  
RC STRAIN=HF-2;  
MEDLINE=22354719; PubMed=1246555;  
RA Sasaki Y., Ishikawa J., Yamashita A., Oshima K., Kanri T., Furuya K., Yoshino C., Horino A., Shiba T., Sasaki T., Hattori M.,  
RT "The complete genomic sequence of Mycoplasma penetrans, an intracellular bacterial pathogen in humans."  
RL Nucleic Acids Res. 30:5293-5300(2002).  
DR EMBL; AP004174; BAC44812.1; -  
KW Complete proteome.  
SQ SEQUENCE 165 AA; 19038 MW; C6851C7B8FFBC54 CRC64;

Query Match 24.1%; Score 194; DB 16; Length 165;  
Best Local Similarity 31.9%; Pred. No. 2.7e-10;  
Matches 51; Conservative 36; Mismatches 57; Indels 16; Gaps 5;

QY 3 RVGMGYDVHRENDGHIILGVKIPYEKGLAHSDDGVVLAADALIGAAALDIGHK 61  
DB KTGNGIDHKKKKKKKQRLAGIDELDKITAHSDGDIILHSISALIGALSLDGLGY 76  
QY 62 PDDDPFKADSRVLLRHVYGVYKKGKYLVAADVTTIIAQAPKMLPHV--PGMRANIA 119  
DB FSDTOSKKNGLDLSLEISMCINELKKQDYSISNVDTITCE-----YIFDIDIDQKS 130  
QY 120 DIEDVD--FINVKAATTEKLGFEGRKEGIAVQAVLIER 157  
DB 131 NLEKLNTEISLKAIR-----YEEDKNMIQVTVALLINK 165

## RESULT 13

Q8XK06 PRELIMINARY; PRT; 89 AA.

AC Q8XK06; 01-OCT-2002 (TREMBlrel. 22, Created)  
DT 01-OCT-2002 (TREMBlrel. 22, Last sequence update)  
DE 2C-methyl-D-erythritol 2,4-cyclodiphosphate synthase (Fragment).  
GN SIR5142.  
OS Synecchococcus sp. (strain RCC 7002) (Agmenellum quadruplicatum).  
OC Bacteria; Cyanobacteria; Chroococcales; Synecchococcus.

OX NCBI\_TaxID=32049;

RN [1]  
RP SEQUENCE FROM N.A.

RA Nomura C.T., Persson S., Zhao J., Bryant D.A.;  
RT An Analysis of Forty Genes Encoding Electron Transport Proteins from  
RT Synecchococcus sp. PCC 7002: A Comparative Study of Electron Transport  
RT Proteins from Cyanobacteria and Chloroplasts."  
RL Submitted (May-2001) to the EMBL/Genbank/DBJ databases.  
DR EMBL; AF381048; AAN03587.1; -  
DR InterPro; IPR003526; YgBB.  
DR Pfam; PF02542; YgBB; 1.  
DR TIGRPFAMs; TIGR00151; 1SPF; 1.  
FT NON TER  
SQ SEQUENCE 89 AA; 9626 MW; 7E5534DC1BF1CE09 CRC64;

Query Match 22.3%; Score 180; DB 2; Length 89;  
Best Local Similarity 43.0%; Pred. No. 2.4e-09;  
Matches 37; Conservative 18; Mismatches 31; Indels 0; Gaps 0;

QY 70 KGADSRVLLRHVYGVYKKGKYLVAADVTTIIAQAPKMLPHVGMKANIADLETDVDFIN 129  
DB 1 KGADSRVLLRHVYGVYKKGKYLVAADVTTIIAQAPKMLPHVGMKANIADLETDVDFIN 129  
QY 130 YKATTEKLGFEGRKEGIAVQAVLI 155  
DB 61 YKATTEKLGFEGRKEGIAVQAVLI 155

## RESULT 14

Q8EFR7 PRELIMINARY; PRT; 571 AA.

AC Q8EFR7; 01-MAR-2003 (TREMBlrel. 23, Created)  
DT 01-MAR-2003 (TREMBlrel. 23, Last sequence update)  
DE Conserved hypothetical protein.  
GN S01899.  
OS Shewanella oneidensis.  
OC Bacteria; Proteobacteria; Gammaproteobacteria; Alteromonadales;  
OC Alteromonadaceae; Shewanella.  
OX NCBI\_TaxID=70863;  
RN [1]  
RP SEQUENCE FROM N.A.  
RC STRAIN=NR-1;  
MEDLINE=22297686; PubMed=12368813;  
RA Heidelberg J.F., Paulsen I.T., Nelson K.E., Gaidos E.J., Nelson M.C.,  
RA Read T.D., Eisen J.A., Seebadri R., Ward N., Meche B., Clayton R.A.,  
RA Meyer T., Tsapin A., Scott J., Beanan M., Brinkac L., Daugherty S.,  
RA Debey R.T., Dodson R.J., Durkin A.S., Haft D.H., Kolonay J.F.,  
RA Madupu R., Peterson J.D., Umeyam L.A., White O., Wolf A.M.,  
RA Vamathevan J., Weidman J., Impraim M., Lee K., Berry K., Lee C.,  
RA Mueller U., Khouli H., Gill J., Utechtack T.R., McDonald L.A.,  
RA Feldblyum T.V., Smith H.O., Venter J.C., Neilson K.H., Fraser C.M.,  
RT "Genome sequence of the dissimilatory metal ion-reducing bacterium  
RT Shewanella oneidensis."  
RL Nat. Biotechnol. 20:1118-1123(2002).  
DR EMBL; AE015633; AANS4951.1; -  
DR TIGR; S01899; -  
KW Hypothetical protein; Complete proteome.  
SQ SEQUENCE 571 AA; 64713 MW; 86E9246E7680D81E CRC64;

Query Match 10.9%; Score 87.5; DB 16; Length 571;  
Best Local Similarity 24.6%; Pred. No. 12;  
Matches 35; Conservative 17; Mismatches 39; Indels 51; Gaps 6;

QY 8 YOVERPNDGHIILGVKIPYEKGLAHSDDGVVLAADALIGAAALDIGHK 54  
DB 141 YAHILNGKVPNSMLGKTYVD--EAHLDFDTLKQLESHIKAHVAAINGLAPKITP 197  
QY 55 -----LSDIGKHF-----DTPDFKADSRVLLRHVYGVYKKGKYL 91  
DB 198 YHOLKQYLAHYKDLAARYPTTITPTVEVIRPGTTSPIKIGIAR-----LKEGGL 248

QY 92 LVNA---DVTIIAAPPMLPHV 110  
 Db 249 DANATVQDRLTVANAPLIDNT 270

## RESULT 15

Q8LHV9 PRELIMINARY; PRT; 434 AA.  
 ID Q8LHV9  
 AC Q8LHV9;  
 DT 01-OCT-2002 (TrEMBLrel. 22, Created)  
 DT 01-OCT-2002 (TrEMBLrel. 22, Last sequence update)  
 DT 01-MAR-2003 (TrEMBLrel. 23, Last annotation update)  
 DE Monodehydroascorbate reductase (NADH)-like protein.  
 OS Arabidopsis thaliana (Mouse-ear cress).  
 OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;  
 OC Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; Rosidae;  
 OC eurosid 11; Brassicales; Brassicaceae; Arabidopsis.  
 NCBI\_taxid=3702;  
 [1]  
 RN SEQUENCE FROM N.A.  
 RP Haas B.J., Vollovsky N., Town C.D., Troughan M., Alexandrov N.,  
 RA Feldmann K.A., Flavell R.B., White O., Salzberg S.L.;  
 RT "Full-length messenger RNA sequences greatly improve genome  
 RL annotation.";  
 genome Biol. 0:0-0(2002).  
 [2]  
 RN SEQUENCE FROM N.A.  
 RP Brover V., Troughan M., Alexandrov N., Lu Y.-P., Flavell R.,  
 RA Feldmann K.;  
 RT "Full-length cDNA from Arabidopsis thaliana.";  
 Submitted (MAR-2002) to the EMBL/GenBank/DBJ databases.  
 CC -1- COFACTOR: FAD (BY SIMILARITY).  
 DR EMBL; AY086968; AAM64531.1; -;  
 DR InterPro; IPR001327; FAD\_Pyr\_redox.  
 DR InterPro; IPR001003; Pyridine\_redox\_2.  
 DR InterPro; IPR001100; Pyr\_redox.  
 DR Pfam; PF00070; Pyr\_redox; 1.  
 DR PRINTS; PR00368; FADPR.  
 DR PRINTS; PR00411; PYRIDTASE1.  
 DR PRINTS; PR00469; PNDPDTASE1.  
 DR PRODOM; PD000139; FAD\_Pyr\_redox; 1.  
 DR FAD; Flavoprotein; Oxidoreductase.  
 SQ SEQUENCE 434 AA; 46486 MW; 0DB0C3PFG81D4C0C CRC64;

Query Match 10.7%; Score 86; DB 10; Length 434;  
 Best Local Similarity 26.7%; Pred. No. 11;  
 Matches 50; Conservative 22; Mismatches 59; Indels 56; Gaps 8;

QY 22 GGVXIPYE-----RGLEAHSDGVVLAHLDALIGAAALDIGH-----PPPTD 66  
 Db 73 GGEKLPESYKQKIEILISTEIVKADLSKSL-VSATGVDFKQYQTLITATGSTVLRITD 131  
 QY 67 PNFKGADSRVLT-----RHVGIIVEKGYKLV-----NADVTI 99  
 Db 132 FGVGKADSKNIILVLRIDADKLVBAIKAKKGGKAVVGGYIGELSAVLHNNLDVTM 191  
 QY 100 IAAQPKLPEVPGKRNIAADLTVDYFNKATL-TEKLGEGR-----KEGTA 148  
 Db 192 VFPEPFCMPEL--FTANIAFETLYTNKGKLIKGTVAAGFTAQPNGEVKEVQLMDGRT 249  
 QY 149 VQAVVLI 155  
 Db 250 LEADIVI 256

Search completed: January 29, 2004, 15:54:10  
 Job time : 13.7026 secs



DR N-PSDB; ABE83264.

XX New high growth methanotrophic bacterial strain, useful for producing  
PT single cell proteins, grows on a C1 carbon substrate, and comprises a  
XX functional gene encoding in Embden-Meyerhof carbon pathway

PS Claim 11; Page 145-147; 157pp; English.

CC The invention relates to a high growth methanotrophic bacterial strain,  
CC which grows on a C1 carbon substrate e.g. methane and methanol, and  
CC comprises a functional Embden-Meyerhof carbon flux pathway comprising a  
CC gene coding a pyrophosphate dependent phosphotransferase enzyme or a 16S  
CC RNA. The bacterial strain is useful for the production of single cell  
CC protein and for the biotransformation of a nitrogen-containing compound,  
CC e.g. ammonia, nitrate, nitrite or nitrogen. It is also useful for the  
CC production of a feed product comprising a protein, carbohydrates and a  
CC pigment and for reducing oxygen demand, for removing nitrates and  
CC nitrates in methane-containing environments such as landfill, waste  
CC water treatment systems or anywhere that methane, oxygen and nitrates are  
CC present. The bacterial strain of the invention can be used as a  
CC denitrifying agent for the conversion of nitrate or nitrite to nitrous  
CC oxide with methane or methanol as a carbon source. It is also used in the  
CC production of biomass including proteins, carbohydrates and a wide  
CC variety of pigments (particularly for isoprenoid pigments for the  
CC purpose of generating animal feeds), in production of terpenoid and  
CC carotenoid compounds, useful as pigments and as monomers in polymeric  
CC materials and in production of exopolysaccharides at high levels.  
CC Sequences ABE61551-ABE61590 represent high growth methanotrophic  
CC bacterial strain proteins of the invention.

XX Sequence 544 AA;

Query Match 100.0%; Score 2788; DB 23; Length 544;  
Best Local Similarity 100.0%; Pred. No. 3e-248;  
Matches 544; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 MTKFPIFTGGVSSISGKGIASSLAAILLEDRGLKATITKLPYINVDGTMSPFGHGVF 60  
DB 1 MTKFPIFTGGVSSISGKGIASSLAAILLEDRGLKATITKLPYINVDGTMSPFGHGVF 60  
QY 61 VTEDEGAEITDLDGHERFLKTTMTKKNFTTQGVYEQVLRNKKGDYLGATVQVPIPHND 120  
DB 61 VTEDEGAEITDLDGHERFLKTTMTKKNFTTQGVYEQVLRNKKGDYLGATVQVPIPHND 120  
QY 121 EIKRRVYESAGQVALIEVGTVGDIESTPLETTRQKGVLEGRRAIFILHTLVPIYK 180  
DB 121 EIKRRVYESAGQVALIEVGTVGDIESTPLETTRQKGVLEGRRAIFILHTLVPIYK 180  
QY 181 SAGELEKTPQHSVKELRTIGIOPDLICRSEOPIPASERRKIALFTVAERKAVISADIA 240  
DB 181 SAGELEKTPQHSVKELRTIGIOPDLICRSEOPIPASERRKIALFTVAERKAVISADIA 240  
QY 241 DTYRIPLLIREQGLDLDVDDLRPAADLSAMERKVDGLTHPTDEVSAIVGRYVDH 300  
DB 241 DTYRIPLLIREQGLDLDVDDLRPAADLSAMERKVDGLTHPTDEVSAIVGRYVDH 300  
QY 301 TDYVKSINELTHAGITHRHKVOISYIDSETISABGTAKIKANDALIVPGSGERGVGRK 360  
DB 301 TDYVKSINELTHAGITHRHKVOISYIDSETISABGTAKIKANDALIVPGSGERGVGRK 360  
QY 361 ISTVFARENKIPYLGICLMOGSAVIEFARNVGLBGASHTEFLPKSPFVIGLITENMD 420  
DB 361 ISTVFARENKIPYLGICLMOGSAVIEFARNVGLBGASHTEFLPKSPFVIGLITENMD 420  
QY 421 EAGEIVTRDSDSLGGMRLGAOKCKLKADSLAPOLYQKQVITERRRHRYEFNNQTLKOL 480  
DB 421 EAGEIVTRDSDSLGGMRLGAOKCKLKADSLAPOLYQKQVITERRRHRYEFNNQTLKOL 480  
QY 481 EAAGKFSKGSIDGRVLEIIELEPHWFLACQPHPEPTSTPRNGHLLFGGFVNAAKHNT 540  
DB 481 EAAGKFSKGSIDGRVLEIIELEPHWFLACQPHPEPTSTPRNGHLLFGGFVNAAKHNT 540  
QY 541 QGTA 544

DB 541 QGTA 544

RESULT 2  
ID ABE22306  
XX ABE22306 standard; Protein; 544 AA.

AC ABE22306;

DT 25-JUL-2002 (First entry)

DB Methylomonas 16a sp. CTP synthase (PyrG) enzyme.

XX Carotenoid; isopentenyl pyrophosphate; anthraxanthin; astaxanthin; diet;

KW anti-oxidant; steroid; flavour; fragrance; electro-optic application;

XX aquaculture; enzyme; CTP synthase; Pyrg.

OS Methylomonas 16a sp.

PD MO200218617-A2.

PF 04-SEP-2001; 2001WO-US27420.

PR 01-SEP-2000; 2000US-223958P.

PR 01-SEP-2000; 2000US-223907P.

XX (DUPO) DU POINT DE MEMOIRS & CO E. I.

PI Brzostowicz PC, Cheng Q, Dicosimo DJ, Koffas M, Miller ES;

XX Odum JM, Picataggio SK, Rouviere PE;

XX MPI; 2002-351711/38.

DR N-PSDB; ABE235504.

PT Producing carotenoid compounds e.g. anthraxanthin and astaxanthin, by

PT using microorganisms having a nucleic acid molecule encoding enzymes in

PT the carotenoid biosynthetic pathway and which metabolize single carbon

PT substrates

PS Claim 46; Page 120-122; 156pp; English.

XX The invention relates to a method for producing carotenoid compounds.

CC The method comprises a transformed metabolizing host cell, comprising

CC suitable levels of isopentenyl pyrophosphate and a nucleic acid molecule

CC encoding an enzyme in the carotenoid biosynthetic pathway, under the

CC control of regulatory sequences, and contacting the host cell with carbon

CC substrate to produce a carotenoid compound. The method is useful for

CC producing carotenoid compounds such as anthraxanthin and astaxanthin, by

CC using microorganisms having a nucleic acid molecule encoding enzymes in

CC the carotenoid biosynthetic pathway and which metabolize single carbon

CC substrates. The carotenoids have potent anti-oxidant properties useful in

CC diet, and aquaculture elements. The carotenoids are also useful as

CC intermediates in the synthesis of steroids flavours and fragrances and

CC compounds for potential electro-optic applications. The present sequence

CC is Methylomonas 16a sp. CTP synthase (PyrG) enzyme used in the invention.

XX Sequence 544 AA;

Query Match 100.0%; Score 2788; DB 23; Length 544;  
Best Local Similarity 100.0%; Pred. No. 3e-248;  
Matches 544; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 MTKFPIFTGGVSSISGKGIASSLAAILLEDRGLKATITKLPYINVDGTMSPFGHGVF 60  
DB 1 MTKFPIFTGGVSSISGKGIASSLAAILLEDRGLKATITKLPYINVDGTMSPFGHGVF 60  
QY 61 VTEDEGAEITDLDGHERFLKTTMTKKNFTTQGVYEQVLRNKKGDYLGATVQVPIPHND 120  
DB 61 VTEDEGAEITDLDGHERFLKTTMTKKNFTTQGVYEQVLRNKKGDYLGATVQVPIPHND 120

QY 121 EIKRRVYESAEKGDVALIEVGGTVGDIESLPFLSTIRQMGVGLGRDRLFIHLTLVPYIK 180  
 DB 121 EIKRRVYESAEKGDVALIEVGGTVGDIESLPFLSTIRQMGVGLGRDRLFIHLTLVPYIK 180  
 QY 181 SAGELKTKPTQHSYKELRTTIGIOPDIIICRSEQPIPAERKIALFTNVAEKAVISIDA 240  
 DB 181 SAGELKTKPTQHSYKELRTTIGIOPDIIICRSEQPIPAERKIALFTNVAEKAVISIDA 240  
 QY 241 DTIYRIPLILREOGDLVDVQDLRLDVPADLSAMEKVVQGLTPRDEVSIAIVGKYVDH 300  
 DB 241 DTIYRIPLILREOGDLVDVQDLRLDVPADLSAMEKVVQGLTPRDEVSIAIVGKYVDH 300  
 QY 301 TDVAKSLNEALIHAGIHTHRKQVSIYIDSETTEABGTAKLKNVDAIIVPGGGERGVEGK 360  
 DB 301 TDVAKSLNEALIHAGIHTHRKQVSIYIDSETTEABGTAKLKNVDAIIVPGGGERGVEGK 360  
 QY 361 ISTVRPAERKIPYGLICLQKQSAVIEFARNVGLBGHSTFELPKSPHVIQGLITEMND 420  
 DB 361 ISTVRPAERKIPYGLICLQKQSAVIEFARNVGLBGHSTFELPKSPHVIQGLITEMND 420  
 QY 421 EAGELVTRDESDSLGTMRLGAQKRLKADSLAFQLYOKDVITERHRRHYEFNNQYLKOL 480  
 DB 421 EAGELVTRDESDSLGTMRLGAQKRLKADSLAFQLYOKDVITERHRRHYEFNNQYLKOL 480  
 QY 481 EAAQMKFSGKSLDGLVLEIILPEHPWFLACQHPERTSTPRNGHALPSGFVEAAAKHKT 540  
 DB 481 EAAQMKFSGKSLDGLVLEIILPEHPWFLACQHPERTSTPRNGHALPSGFVEAAAKHKT 540  
 QY 541 QGTA 544  
 DB 541 QGTA 544

RESULT 3  
 AAU80330  
 ID AAU80330 standard; Protein: 544 AA.  
 XX  
 AC AAU80330;  
 XX  
 DT 15-JUL-2002 (first entry)  
 XX  
 DE Methylomonas 16a ORF6 pyrg (CTP synthase enzyme) protein sequence.  
 XX  
 KM Isoprenoid biosynthetic enzyme; isoprenoid compound; feed additive;  
 KM keratenoic; pigment; flavour; fragrance; open reading frame 6; ORF6;  
 KM pyrg; CTP synthase enzyme.  
 XX  
 OS Methylomonas sp.  
 XX  
 PN WC0200220733-A2.  
 XX  
 PD 14-MAR-2002.  
 XX  
 PF 29-AUG-2001; 2001MO-US26852.  
 XX  
 PR 01-SEP-2000; 2000US-229907P.  
 XX  
 RA (DUPO) DU PONT DE NEMOURS & CO E I.  
 XX  
 PI Cheng Q, Kofas M, Norton KC, Odom JM, Picataggio SK, Rouviere PE;  
 PI Schenzle A, Tomb J;  
 XX  
 DR WPI; 2002-383051/41.  
 DR N-PSDB; ABK50086.  
 XX  
 PT Novel nucleic acid molecule encoding a isoprenoid biosynthetic enzyme,  
 PT isolated from Methylomonas 16a, useful for the production of isoprenoid  
 compounds -  
 XX  
 PS Claim 4; Page 74-76; 84pp; English.  
 CC The present invention relates to a new nucleic acid molecule encoding  
 CC an isoprenoid biosynthetic enzyme isolated from Methylomonas 16a.

CC The invention is useful for obtaining a nucleic acid molecule  
 CC encoding an isoprenoid compound biosynthetic enzyme, and for the  
 CC microbial production of isoprenoid compounds. The molecules of the  
 CC invention are also useful for regulating isoprenoid biosynthesis in an  
 CC organism and for producing recombinant organisms for producing various  
 CC isoprenoid compounds. The nucleic acid is also useful for feed additive,  
 CC for the production of keratenoic and their derivatives, isoprenoid  
 CC intermediates, and as pure products useful as pigments, flavours and  
 CC fragrances. The present amino acid sequence represents the Methylomonas  
 CC 16a open reading frame 6 (ORF6) pyrg (CTP synthase enzyme) protein of  
 CC the invention, as described above.  
 XX  
 Sequence - 544 AA:  
 Query Match 100.0%; Score 2788; DB 23; Length 544;  
 Best local Similarity 100.0%; Pred. No. 3e-248;  
 Matches 544; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
 QY 1 MTKFIFITGGVSSLSKGIAASSLAILEDRLKVTITLADPYINVPQTMSPFGHGVF 60  
 DB 1 MTKFIFITGGVSSLSKGIAASSLAILEDRLKVTITLADPYINVPQTMSPFGHGVF 60  
 QY 61 VTEKCASTYLDLGHYERFLKTTTKNNFTTQVYEOVLRNKRKDYLGATVQVPIPHID 120  
 DB 61 VTEKCASTYLDLGHYERFLKTTTKNNFTTQVYEOVLRNKRKDYLGATVQVPIPHID 120  
 QY 121 EIKRRVYESAEKGDVALIEVGGTVGDIESLPFLSTIRQMGVGLGRDRLFIHLTLVPYIK 180  
 DB 121 EIKRRVYESAEKGDVALIEVGGTVGDIESLPFLSTIRQMGVGLGRDRLFIHLTLVPYIK 180  
 QY 181 SAGELKTKPTQHSYKELRTTIGIOPDIIICRSEQPIPAERKIALFTNVAEKAVISIDA 240  
 DB 181 SAGELKTKPTQHSYKELRTTIGIOPDIIICRSEQPIPAERKIALFTNVAEKAVISIDA 240  
 QY 241 DTIYRIPLILREOGDLVDVQDLRLDVPADLSAMEKVVQGLTPRDEVSIAIVGKYVDH 300  
 DB 241 DTIYRIPLILREOGDLVDVQDLRLDVPADLSAMEKVVQGLTPRDEVSIAIVGKYVDH 300  
 QY 301 TDVAKSLNEALIHAGIHTHRKQVSIYIDSETTEABGTAKLKNVDAIIVPGGGERGVEGK 360  
 DB 301 TDVAKSLNEALIHAGIHTHRKQVSIYIDSETTEABGTAKLKNVDAIIVPGGGERGVEGK 360  
 QY 361 ISTVRPAERKIPYGLICLQKQSAVIEFARNVGLBGHSTFELPKSPHVIQGLITEMND 420  
 DB 361 ISTVRPAERKIPYGLICLQKQSAVIEFARNVGLBGHSTFELPKSPHVIQGLITEMND 420  
 QY 421 EAGELVTRDESDSLGTMRLGAQKRLKADSLAFQLYOKDVITERHRRHYEFNNQYLKOL 480  
 DB 421 EAGELVTRDESDSLGTMRLGAQKRLKADSLAFQLYOKDVITERHRRHYEFNNQYLKOL 480  
 QY 481 EAAQMKFSGKSLDGLVLEIILPEHPWFLACQHPERTSTPRNGHALPSGFVEAAAKHKT 540  
 DB 481 EAAQMKFSGKSLDGLVLEIILPEHPWFLACQHPERTSTPRNGHALPSGFVEAAAKHKT 540  
 QY 541 QGTA 544  
 DB 541 QGTA 544

RESULT 4  
 AAU36340  
 ID AAU36340 standard; Protein: 542 AA.  
 XX  
 AC AAU36340;  
 XX  
 DT 14-FEB-2002 (first entry)  
 XX  
 DE Pseudomonas aeruginosa cellular proliferation protein #330.  
 XX  
 DR Antisense; prokaryotic cellular proliferation protein;  
 XX antibiotic; antibacterial; drug design.  
 CC Pseudomonas aeruginosa.

```

XX  NO2001:70955-A2.
XX  27-SEP-2001.
XX  21-MAR-2001; 2001WO-US09180.
XX  21-MAR-2000; 2000US-191078P.
XX  23-MAY-2000; 2000US-206848P.
XX  26-MAY-2000; 2000US-207727P.
XX  27-OCT-2000; 2000US-242578P.
XX  27-NOV-2000; 2000US-253625P.
XX  22-DEC-2000; 2000US-257931P.
XX  16-FEB-2001; 2001US-269308P.
XX  (ELITR-) ELITRA PHARM INC.
XX  Haselbeck R, Ohlsen KL, Zyskind JW, Wall D, Trawick JD, Carr GU,
XX  Yamamoto RT, Xu HH;
XX  WPI; 2001-611495/70.
XX  N-PSDB; AAS54199.
XX  New polynucleotides for the identification and development of
XX  antibiotics, comprise sequences of antisense nucleic acids -
XX  Example 3; Seq ID No 11933; 511pp; English.
XX  The invention relates to antisense inhibitors of genes essential to
XX  prokaryotic cellular proliferation, their use in identifying the
XX  genes, their use in the discovery of novel antibiotics, the essential
XX  genes themselves and the encoded proteins. The prokaryotes used are
XX  Escherichia coli, Staphylococcus aureus, Salmonella typhi, Klebsiella
XX  pneumoniae, Pseudomonas aeruginosa and Enterococcus faecalis. The
XX  invention is also useful for the identification of potential new targets
XX  for antibiotic development. The antisense nucleic acids can also be used
XX  to identify proteins used in proliferation, to express these proteins,
XX  and to obtain antibodies capable of binding to the expressed proteins.
XX  The proteins can be used to screen compounds in rational drug discovery
XX  programmes. The antisense nucleic acid sequence is also useful to screen
XX  for homologous nucleic acids which are required for cell proliferation in
XX  a wide variety of organisms. The present sequence represents an
XX  essential prokaryotic cellular proliferation protein.
XX  Note: The sequence data for this patent did not form part
XX  of the printed specification, but was obtained in electronic
XX  format directly from WIPO at
XX  ftp.wipo.int/pub/published_pct_sequences.
XX  Sequence 542 AA.
XX  Query Match 72.0%; Score 2008; DB 22; Length 542;
XX  Best Local Similarity 71.2%; Pred. No. 3e-176;
XX  Matches 383; Conservative 64; Mismatches 91; Indels 0; Gaps 0;
XX  1 MTKEFTTGVSSLSKGLAASLAALIEDRGKVTITLDPYINVDPGTMSFGHGEVF 60
XX  1 MTKYIVTGVSSLSKGLAASLAALIEDRGKVTITLDPYINVDPGTMSFGHGEVF 60
XX  1 MTKYIVTGVSSLSKGLAASLAALIEDRGKVTITLDPYINVDPGTMSFGHGEVF 60
XX  61 VTEDGAEFDLDLGHYERFLKTTTKKNNFTTGVYEQVLNENKGDYLGATYQVPIPHITD 120
XX  61 VTEDGAEFDLDLGHYERFLKTTTKKNNFTTGVYEQVLNENKGDYLGATYQVPIPHITD 120
XX  61 VTODGAEFDLDLGHYERFLKTTTKKNNFTTGVYEQVLNENKGDYLGATYQVPIPHITD 120
XX  121 EIKRRVYSEAGKDVALLIIVGTVGDIIEIPELETTRKMGVDELGRRAFLHLLTVPIYK 180
XX  121 EIKRRVYSEAGKDVALLIIVGTVGDIIEIPELETTRKMGVDELGRRAFLHLLTVPIYK 180
XX  121 EIKRRVYSEAGKDVALLIIVGTVGDIIEIPELETTRKMGVDELGRRAFLHLLTVPIYK 180
XX  181 SAGELTKTQHSVKELRTIGIOPDILICRSEQPIPASRRKIALFTYNAEKVISAIDA 240
XX  181 SAGELTKTQHSVKELRTIGIOPDILICRSEQPIPASRRKIALFTYNAEKVISAIDA 240
XX  181 TAGETKTKPQHSVKELRTIGIOPDILICRSEQPIPASRRKIALFTYNAEKVISAIDA 240
XX  241 DTIYRIPILREGGLDLYVDQRLDVPADLSAMKRYVDGLTHPIDEYSINAVKRYVDH 300
XX  241 DTIYRIPILREGGLDLYVDQRLDVPADLSAMKRYVDGLTHPIDEYSINAVKRYVDH 300
XX  241 DTIYRIPILREGGLDLYVDQRLDVPADLSAMKRYVDGLTHPIDEYSINAVKRYVDH 300

```

```

XX  301 TDAVSLINELIAGIHTRHRVQISYIDSETEAGTAKLNKVDAILVGGGNGEVCK 360
XX  301 TDAVSLINELIAGIHTRHRVQISYIDSETEAGTAKLNKVDAILVGGGNGEVCK 360
XX  301 EDVYSLIEMTHAGIOSRTKYNLRIEDIEQGTILBVDAILVGGGNGEVCK 360
XX  361 ISTVAPARENKIPYIGICLQMSAVIEFARNVVGEGASHTFEFLKSPHPVGLITEMD 420
XX  361 ISTVAPARENKIPYIGICLQMSAVIEFARNVVGEGASHTFEFLKSPHPVGLITEMD 420
XX  361 ISTVAPARENKIPYIGICLQMSAVIEFARNVVGEGASHTFEFLKSPHPVGLITEMD 420
XX  421 EAGELVTRDESDLDGQVRLGAKQKRLADSLAFLQYKQDVITERRHRHRYEFNNYKQL 480
XX  421 EAGELVTRDESDLDGQVRLGAKQKRLADSLAFLQYKQDVITERRHRHRYEFNNYKQL 480
XX  421 ATGATEIRTEAGDAGTWRIGAGQECQLGTGLVHDCYAKDYLVERHRRRYEYNNMLPOL 480
XX  481 EAGMKTFGSKSLDGLVLIIELEPHWPVLACQHPHETSTPRNKALTSQFVEAAKH 538
XX  481 EAGMKTFGSKSLDGLVLIIELEPHWPVLACQHPHETSTPRNKALTSQFVEAAKH 538
XX  481 EAGMKTFGSKSLDGLVLIIELEPHWPVLACQHPHETSTPRNKALTSQFVEAAKH 538
XX  RESULT 5
XX  AAU34682
XX  ID AAU34682 standard; Protein; 545 AA.
XX  AAU34682;
XX  14-FEB-2002 (first entry)
XX  E. coli cellular proliferation protein #263.
XX  Antisense; prokaryotic cellular proliferation protein;
XX  antibiotic; antibacterial; drug design.
XX  Escherichia coli.
XX  WO200170955-A2.
XX  27-SEP-2001.
XX  21-MAR-2001; 2001WO-US09180.
XX  21-MAR-2000; 2000US-191078P.
XX  23-MAY-2000; 2000US-206848P.
XX  26-MAY-2000; 2000US-207727P.
XX  27-OCT-2000; 2000US-242578P.
XX  27-NOV-2000; 2000US-253625P.
XX  22-DEC-2000; 2000US-257931P.
XX  16-FEB-2001; 2001US-269308P.
XX  (ELITR-) ELITRA PHARM INC.
XX  Haselbeck R, Ohlsen KL, Zyskind JW, Wall D, Trawick JD, Carr GU,
XX  Yamamoto RT, Xu HH;
XX  WPI; 2001-611495/70.
XX  N-PSDB; AAS52541.
XX  New polynucleotides for the identification and development of
XX  antibiotics, comprise sequences of antisense nucleic acids -
XX  Example 3; Seq ID No 10275; 511pp; English.
XX  The invention relates to antisense inhibitors of genes essential to
XX  prokaryotic cellular proliferation, their use in identifying the
XX  genes, their use in the discovery of novel antibiotics, the essential
XX  genes themselves and the encoded proteins. The prokaryotes used are
XX  Escherichia coli, Staphylococcus aureus, Salmonella typhi, Klebsiella
XX  pneumoniae, Pseudomonas aeruginosa and Enterococcus faecalis. The
XX  invention is also useful for the identification of potential new targets
XX  for antibiotic development. The antisense nucleic acids can also be used
XX  to identify proteins used in proliferation, to express these proteins,
XX  and to obtain antibodies capable of binding to the expressed proteins.
XX  The proteins can be used to screen compounds in rational drug discovery
XX  programmes. The antisense nucleic acid sequence is also useful to screen
XX  for homologous nucleic acids which are required for cell proliferation in

```

CC a wide variety of organisms. The present sequence represents an  
 CC essential prokaryotic cellular proliferation protein.  
 CC Note: The sequence data for this patent did not form part  
 CC of the printed specification, but was obtained in electronic  
 CC format directly from WIPO at  
 CC ftp.wipo.int/pub/published\_pct\_sequences.

XX Sequence 545 AA;

Query Match 68.7%; Score 1916; DB 22; Length 545;  
 Best Local Similarity 67.7%; Pred. No. 9.3e-168;

Matches 363; Conservative 71; Mismatches 102; Indels 0; Gaps 0;

```

QY 2 TKFIFITGGVSSSLGKGIASSLAALIEDRGKLYTTKLDPIYNDPGTMSPTQGEVAV 61
QY 3 TNYIFVTGGVSSSLGKGIASSLAALIEDRGKLYTTKLDPIYNDPGTMSPTQGEVAV 62
DB 62 TEDGAEITDLDLGHYERFLTKTMTKNNFTTGVYEQVLRNKKGYLGATVOYIPIHIDE 121
DB 63 TEDGAEITDLDLGHYERFLTKTMTKNNFTTGVYEQVLRNKKGYLGATVOYIPIHIDE 122
QY 122 IKRRVYSABGKQVALIEVGGTVDIESLPLETITQMGVEIGRDLAFIHLTLVPYIKS 181
DB 123 IKRRVYSABGKQVALIEVGGTVDIESLPLETITQMGVEIGRDLAFIHLTLVPYIKS 182
QY 182 AGSLKTPITQHSYKELRTIGIQPDILICRSEOPIPASERRKIALFTNAEKAVISAIDAD 241
DB 183 SGSEVTKPTQHSYKELRTIGIQPDILICRSDRAVPANERAKIALFCNVEPKAVISLKQVD 242
QY 242 TIRIRPLLRGQGLDVLVDQLADLPADLSAMKENVGDLTHPTDEVSIATVYKRVYHT 301
DB 243 STIKIRPLLRGQGLDVLVDQLADLPADLSAMKENVGDLTHPTDEVSIATVYKRVYHT 302
QY 302 DAYKSINLALIHAGIHTHRAVOISYIDSETIEAGTAKLVNDAILVPGFGERGVGSKI 361
DB 303 DAYKSINLALIHAGIHTHRAVOISYIDSETIEAGTAKLVNDAILVPGFGERGVGSKI 362
QY 362 STYRPARBNKIPLYGICLGMQSAVIEFARNVGLBGASTFELPSRPHVIGLITEMWDE 421
DB 363 STYRPARBNKIPLYGICLGMQSAVIEFARNVGLBGASTFELPSRPHVIGLITEMWDE 422
QY 422 AGELVTRDESDSLGGMRLGAOKCRLKASLAFOLQOKVITERHHRERFENNQVYKOLE 481
DB 423 NGWVTRSESDSLGGMRLGAOKCRLKASLAFOLQOKVITERHHRERFENNQVYKOLE 482
QY 482 AAGKESGKSLDLRLVYIELPEHPFELACQFPERFTSPRNCHALFSGFVEAAK 537
DB 483 DAGLRVAGSGDDQVLVEIIEVPHNPVACQFPERFTSPRNCHALFSGFVEAAK 538

```

# RESULT 6

ID AAU35582 standard; Protein; 545 AA.

XX AAU35582;

DT 14-FEB-2002 (first entry)

XX Haemophilus influenzae cellular proliferation protein #223.

XX Antisense; prokaryotic cellular proliferation protein;

XX antibiotic; antibacterial; drug design.

XX Haemophilus influenzae.

XX WO200170955-A2.

XX 27-SEP-2001.

XX 21-MAR-2001; 2001MO-US09180.

XX 21-MAR-2000; 2000US-191078P.

XX 23-MAY-2000; 2000US-206848P.

PR 26-MAY-2000; 2000US-207727P.  
 PR 23-OCT-2000; 2000US-242578P.  
 PR 27-NOV-2000; 2000US-253625P.  
 PR 22-DEC-2000; 2000US-257931P.  
 PR 16-FEB-2001; 2001US-269308P.

PA (ELIT-) ELITRA PHARM INC.

PI Haselbeck R, Ohlsen KU, Zyskind JW, Wall D, Trawick JD, Carr GJ,

PI Yamamoto R, Xu HH;

DR HPI; 2001-611495/70.

DR N-PSDB; AAS53441.

PT New polynucleotides for the identification and development of

PS antibiotics, comprise sequences of antisense nucleic acids -

XX Example 3; Seq ID No 11175; 511pp; English.

The invention relates to antisense inhibitors of genes essential to prokaryotic cellular proliferation, their use in identifying the genes themselves and the encoded proteins. The prokaryotes used are Escherichia coli, Staphylococcus aureus, Salmonella typhi, Klebsiella pneumoniae, Pseudomonas aeruginosa and Enterococcus faecalis. The invention is also useful for the identification of potential new targets for antibiotic development. The antisense nucleic acids can also be used to identify proteins used in proliferation, to express these proteins, and to obtain antibodies capable of binding to the expressed proteins. The proteins can be used to screen compounds in rational drug discovery programs. The antisense nucleic acid sequence is also useful to screen for homologous nucleic acids which are required for cell proliferation in a wide variety of organisms. The present sequence represents an essential prokaryotic cellular proliferation protein.

CC Note: The sequence data for this patent did not form part of the printed specification, but was obtained in electronic format directly from WIPO at

CC ftp.wipo.int/pub/published\_pct\_sequences.

XX Sequence 545 AA;

Query Match 68.5%; Score 1908.5; DB 22; Length 545;

Best Local Similarity 67.5%; Pred. No. 4.6e-167; Indels 1; Gaps 1;

Matches 364; Conservative 71; Mismatches 103; Indels 1; Gaps 1;

```

QY 2 TKFIFITGGVSSSLGKGIASSLAALIEDRGKLYTTKLDPIYNDPGTMSPTQGEVAV 61
QY 3 TNYIFVTGGVSSSLGKGIASSLAALIEDRGKLYTTKLDPIYNDPGTMSPTQGEVAV 62
DB 62 TEDGAEITDLDLGHYERFLTKTMTKNNFTTGVYEQVLRNKKGYLGATVOYIPIHIDE 121
DB 63 TEDGAEITDLDLGHYERFLTKTMTKNNFTTGVYEQVLRNKKGYLGATVOYIPIHIDE 122
QY 122 IKRRVYSABGKQVALIEVGGTVDIESLPLETITQMGVEIGRDLAFIHLTLVPYIKS 181
DB 123 IKRRVYSABGKQVALIEVGGTVDIESLPLETITQMGVEIGRDLAFIHLTLVPYIKS 182
QY 182 AGSLKTPITQHSYKELRTIGIQPDILICRSEOPIPASERRKIALFTNAEKAVISAIDAD 241
DB 183 AGSLKTPITQHSYKELRTIGIQPDILICRSEOPIPASERRKIALFTNAEKAVISAIDAD 242
QY 242 TYRIRPLLRGQGLDVLVDQLADLPADLSAMKENVGDLTHPTDEVSIATVYKRVYHT 301
DB 243 TYRIRPLLRGQGLDVLVDQLADLPADLSAMKENVGDLTHPTDEVSIATVYKRVYHT 302
QY 302 DAYKSINLALIHAGIHTHRAVOISYIDSETIEAGTAKLVNDAILVPGFGERGVGSKI 361
DB 303 DAYKSINLALIHAGIHTHRAVOISYIDSETIEAGTAKLVNDAILVPGFGERGVGSKI 362
QY 362 STYRPARBNKIPLYGICLGMQSAVIEFARNVGLBGASTFELPSRPHVIGLITEMWDE 421
DB 363 STYRPARBNKIPLYGICLGMQSAVIEFARNVGLBGASTFELPSRPHVIGLITEMWDE 422

```

QY 422 AGELVTRDESDLDGGTMRGACQKRLKADSLAFOLYOKDVTTERHRRRYEFNNQYLKOLE 481  
 DB 423 EGNTRVTRDSDDIGTMRGACQCHLVSGSRARELVKERTIEERHRRRYEVNNTLLPOLIE 482  
 QY 482 AAGMKRSGKSLDGLVETIELPEHPWFLACQFHPPEFTSPNGHALPSCVTEAA-AKH 539  
 DB 483 KAGLKVATGLSADKKLVETIEVPHNPFVACQFHPPEFTSPDRGHPFLPAGVKAVERNK 541

## RESULT 7

ABP76969  
 ID ABP76969 standard; Protein; 544 AA.

AC ABP76969;

DT 07-MAR-2003 (first entry)

DE N. gonorrhoeae amino acid sequence SEQ ID 468.

KW Antibacterial; infection; vaccine; gene therapy.

OS Neisseria gonorrhoeae.

PN WO200279243-A2.

PD 10-OCT-2002.

PF 12-FEB-2002; 2002WO-IB02069.

PR 12-FEB-2001; 2001GB-0003424.

XX (CHIR-) CHIRON SPA.

PI Fontana MR, Pizsa M, Masignani V, Monaci E;

DR WPI; 2003-058415/05.

DR N-PSDB; AB237939.

PT New protein from Neisseria gonorrhoeae, useful for the manufacture of a

PT medicament for treating or preventing N. gonorrhoeae infection.

XX Disclosure; Page 222; 815pp; English.

XX The present invention relates to proteins from Neisseria gonorrhoeae.

CC Also disclosed are the nucleic acid molecules encoding the proteins and

CC antibodies that specifically bind to the proteins. The composition

CC comprising the protein, nucleic acid or antibody is useful for the

CC manufacture of a medicament for treating or preventing N. gonorrhoeae

CC infection, this may be in the form of a vaccine or gene therapy.

CC Sequences given in records ABP76736-ABP81046 represent nucleic acid

CC molecules of the invention.

XX Sequence 544 AA;

XX Query Match 67.8%; Score 1889.5; DB 24; Length 544;

XX Best local similarity 66.2%; Pred. No. 2.6e-165;

XX Matches 359; Conservative 78; Mismatches 102; Indels 3; Gaps 2;

QY 1 MTKFIFITGGVSVSLKGIATAASLAILDRGLKXVITTLCDPIYINDPGTMSPFGGEVYF 60  
 DB 1 MTKFIFITGGVSVSLKGIATAASLAILDRGLKXVITTLCDPIYINDPGTMSPFGGEVYF 60  
 QY 61 VTDEGATDLDLGHVERFLKTTWKANNFTTQGVVEQVLRNKRKQDYLDATVQVPIPHITD 120  
 DB 61 VTDEGATDLDLGHVERFLKTTWKANNFTTQGVVEQVLRNKRKQDYLDATVQVPIPHITD 120  
 QY 121 EIKRRVYBSAGKQVALLIEVGTVGDIESTPFTETTRQGVELGRDRAFLPHLTLPYIK 180  
 DB 121 EIKRRVYBSAGKQVALLIEVGTVGDIESTPFTETTRQGVELGRDRAFLPHLTLPYIK 180  
 QY 181 SAGELTKRTPQSVKELRTGIQDPIILICRSBOPIPASERRKIALFTVAEKAIVASIDA 240  
 DB 181 SAGELTKRTPQSVKELRTGIQDPIILICRSBOPIPASERRKIALFTVAEKAIVASIDA 240  
 QY 181 AAGEIKTKPTQIVKMKLSIGLQPDILICEMDRKMKADERRKIALFCVNERAIVGSDIV 240

QY 241 DTVIRIPILLRQGGDDLVQDLRLDVPAPADLSANEXVVDGLTPTDEVSLAIVGKYVDH 300  
 DB 241 DTVIRIPILLRQGGDDLVQDLRLDVPAPADLSANEXVVDGLTPTDEVSLAIVGKYVDH 300  
 QY 301 TDAYSILNEALIHAGIHFTRKVOISYIDSETEAE--GTAKLKNVDAILVPGGEGEVE 358  
 DB 301 TDAYSILNEALIHAGIHFTRKVOISYIDSETEAE--GTAKLKNVDAILVPGGEGEVE 358  
 QY 359 GKISTVFARRENKITYLIGICGMSAVIEFARNVYVLEGASTERTLPSPHVLILTEM 418  
 DB 359 GKISTVFARRENKITYLIGICGMSAVIEFARNVYVLEGASTERTLPSPHVLILTEM 418  
 QY 419 MDEAGELYTRDESDLDGGTMRGACQKRLKADSLAFOLYOKDVTTERHRRRYEFNNQYLK 478  
 DB 419 MDEAGELYTRDESDLDGGTMRGACQKRLKADSLAFOLYOKDVTTERHRRRYEFNNQYLK 478  
 QY 479 QLEAAGKTSFGKSLD-GLVETIELPEHPWFLACQFHPPEFTSPNGHALPSCVTEAAK 537  
 DB 479 QLEAAGKTSFGKSLD-GLVETIELPEHPWFLACQFHPPEFTSPNGHALPSCVTEAAK 537  
 QY 538 HK 539  
 DB 541 NK 542

## RESULT 8

AU38272  
 ID AU38272 standard; Protein; 511 AA.

AC AU38272;

DT 14-FEB-2002 (first entry)

DE Salmonella typhi cellular proliferation protein #163.

KW Antisense; prokaryotic cellular proliferation protein;

KW antibiotic; antibacterial; drug design.

OS Salmonella typhi.

PN WO200170955-A2.

PD 27-SEP-2001.

PF 21-MAR-2001; 2001WO-US09180.

PR 21-MAR-2000; 2000US-191078P.

PR 23-MAY-2000; 2000US-206848P.

PR 26-MAY-2000; 2000US-207727P.

PR 23-OCT-2000; 2000US-242578P.

PR 27-NOV-2000; 2000US-253625P.

PR 22-DEC-2000; 2000US-257931P.

PR 16-FEB-2001; 2001US-269308P.

PA (ELIT-) ELITRA PHARM INC.

PI Haselbeck R, Ohlsen KU, Zyckind JW, Wall D, Trauick JD, Carr GJ;

PI Yamamoto RT, Xu HH;

DR WPI; 2001-611495/70.

DR N-PSDB; AAS56131.

PT New polynucleotides for the identification and development of

PT antibiotics, comprise sequences of antisense nucleic acids -

Example 3; Seq ID No 13865; 511pp; English.

The invention relates to antisense inhibitors of genes essential to prokaryotic cellular proliferation, their use in identifying the genes, their use in the discovery of novel antibiotics, the essential genes themselves and the encoded proteins. The prokaryotes used are *Escherichia coli*, *Staphylococcus aureus*, *Salmonella typhi*, *Klebsiella*

CC pneumoniae, Pseudomonas aeruginosa and Enterococcus faecalis. The  
 CC invention is also useful for the identification of potential new targets  
 CC for antibiotic development. The antisense nucleic acids can also be used  
 CC to identify proteins used in proliferation, to express these proteins,  
 CC and to obtain antibodies capable of binding to the expressed proteins.  
 CC The proteins can be used to screen compounds in rational drug discovery  
 CC programs. The antisense nucleic acid sequence is also useful to screen  
 CC for homologous nucleic acids which are required for cell proliferation in  
 CC a wide variety of organisms. The present sequence represents an  
 CC essential prokaryotic cellular proliferation protein.

CC Note: The sequence data for this patent did not form part  
 CC of the printed specification, but was obtained in electronic  
 CC format directly from WIPO at  
 CC ftp.wipo.int/pub/published\_pct\_sequences.

XX Sequence 511 AA;

Query Match 64.0%; Score 1783; DB 22; Length 511;  
 Best Local Similarity 67.4%; Pred. No. 1.6e-155;  
 Matches 341; Conservative 64; Mismatches 101; Indels 0; Gaps 0;

```

QY 2 TKRFIFINGVSSLSGKGIASSLAILEDRLKVTITKLPYINVDPGTSPQHGGEVF 61
DB 3 TNYIFVIGVSSLSGKGIASSLAILEDRLKVTITKLPYINVDPGTSPQHGGEVF 62
QY 62 TEDGAEITDLGHERLTKMTKKNFTTGOVYEOVLNRRKGDYIGATVOYIPIHTD 121
DB 63 TEDGAEITDLGHERLTKMTKKNFTTGOVYEOVLNRRKGDYIGATVOYIPIHTD 122
QY 122 IKRRVYASAGKVALIEVGTVGDIESLPFLFTIRMGVELGRDALFIHLTLVPIYS 181
DB 123 IKRRVYASAGKVALIEVGTVGDIESLPFLFTIRMGVELGRDALFIHLTLVPIYS 182
QY 182 AGELTKPTQHSVKELETTIGIOPDILCRSEOPIPASERKIALFTNVAEKAVISAI 241
DB 183 AGELTKPTQHSVKELETTIGIOPDILCRSEOPIPASERKIALFTNVAEKAVISAI 242
QY 242 TIYRPLILREOGDLVDQLRLDVPADLSAMEKVVNGLTPTDEVSAIYGVYDHT 301
DB 243 SIYKIFGLTSGDLVDYICRFSINCEANLSEMQIYEEANPAGEVITGMVGYIEHP 302
QY 302 DAKSYNEALIHAGIHTRHKVOISYIDSETTEABGTAKLKNVDAILVPGFGEVGEKI 361
DB 303 DAKSYNEALIHAGIHTRHKVOISYIDSETTEABGTAKLKNVDAILVPGFGEVGEKI 362
QY 362 STVRFARENKIPYIGICLGMOSAVIEFARNVGLGASHTEFLPKSPHYIGLITEMMD 421
DB 363 ATARVARENKIPYIGICLGMOSAVIEFARNVGLGASHTEFLPKSPHYIGLITEMMD 422
QY 422 AGELVTRDESDSLGTMRLGAOKRLKADSLAFOLYOKDYITERHRRYEPNNQYLK 481
DB 423 DGNVAVRSEKSDLGTMRLGAOKRLKADSLAFOLYOKDYITERHRRYEPNNQYLK 482
QY 482 AAGMKFSKSLDLRLVETIELPEHPV 507
DB 483 AAGMKFSKSLDLRLVETIELPEHPV 508

```

RESULT 9  
 AAR15642  
 ID AAR15642 standard; Protein; 535 AA.

XX AAR15642;  
 XX  
 AC AAR15642;  
 XX  
 DT 25-MAR-2003 (updated)  
 DT 05-FEB-1992 (first entry)  
 XX  
 XX CTP synthetase.  
 DE  
 XX Cytidine.  
 XX  
 OS Bacillus natto C-1.  
 XX

PN JP03228689-A.  
 XX  
 XX 09-OCT-1991.  
 PD  
 XX  
 XX 05-FEB-1990; 90JP-0025574.  
 PF  
 XX  
 XX 05-FEB-1990; 90JP-0025574.  
 PR  
 XX  
 XX (ASAH) ASAMI CHEM IND CO LTD.  
 PA  
 XX  
 XX WPI; 1991-343171/47.  
 DR  
 XX N-PSDB; AAQ14668.  
 DR  
 XX

PT Prodn. of cytidine by fermentation with increased yield - comprises  
 PT transforming a suitable microorganism with a recombinant DNA contg.  
 PT CTP synthetase gene and incubating with a uracil cpd.  
 PT  
 XX  
 XX Disclosure; Fig 2; 8pp; Japanese.

CC Microorganisms transformed with the gene can be used to produce  
 CC Cytidine for use as a starting cpd. for prodn. of drugs.  
 CC (Updated on 25-MAR-2003 to correct PA field.)  
 CC

XX Sequence 535 AA;

Query Match 58.2%; Score 1622.5; DB 12; Length 535;  
 Best Local Similarity 57.0%; Pred. No. 1.1e-140;  
 Matches 310; Conservative 94; Mismatches 127; Indels 13; Gaps 3;

```

QY 1 MKRFIFITGVSSLSGKGIASSLAILEDRLKVTITKLPYINVDPGTSPQHGGEVF 60
DB 1 MKRYIFVITGVSSLSGKGIASSLAILEDRLKVTITKLPYINVDPGTSPQHGGEVF 60
QY 61 VEDGAEITDLGHERLTKMTKKNFTTGOVYEOVLNRRKGDYIGATVOYIPIHTD 120
DB 61 VEDGAEITDLGHERLTKMTKKNFTTGOVYEOVLNRRKGDYIGATVOYIPIHTD 120
QY 121 IKRRVYASAG--GNOVALIEVGTVGDIESLPFLFTIRMGVELGRDALFIHLTLV 178
DB 121 IKRRVYASAGKETAADVITTEIGTVGDIESLPFLFTIRMGVELGRDALFIHLTLV 180
QY 179 ITSAGELTKPTQHSVKELETTIGIOPDILCRSEOPIPASERKIALFTNVAEKAVISAI 238
DB 181 IKAAGELTKPTQHSVKELETTIGIOPDILCRSEOPIPASERKIALFTNVAEKAVISAI 240
QY 239 DADTITRIPILREOGDLVDQLRLDVPADLSAMEKVVNGLTPTDEVSAIYGVYDHT 298
DB 241 DADNITRIPILREOGDLVDQLRLDVPADLSAMEKVVNGLTPTDEVSAIYGVYDHT 300
QY 299 DHTDAVKSLNEALIHAGIHTRHKVOISYIDSETTEABGTAKLKNVDAILVPGFGEV 357
DB 301 ELDPDAVYISVVEELRHAGYAFPTDVYVKKWINEAEVTENNIAELTSGDGIYVGGGDRGV 360
QY 358 BCKITVAPARENKIPYIGICLGMOSAVIEFARNVGLGASHTEFLPKSPHYIGLITE 417
DB 361 BCKITVATARENKIPYIGICLGMOSAVIEFARNVGLGASHTEFLPKSPHYIGLITE 420
QY 418 WMDENAGBLVTRDESDSLGTMRLGAOKRLKADSLAFOLYOKDYITERHRRYEPNNQYL 477
DB 421 QMDV-----EDLGGTIRLGLYPCKLBOGTAAFEVYDDEVYVEHRRHRYEPNNFR 470
QY 478 KOLEAAGKKFGSKSLDLRLVETIELPEHPVFLACQHFEEPTSPUNGALTSQVPEAAK 537
DB 471 QOMESQGFVFGTSPDGLVEITELKDPWVAQOFHEFESRPTRPOLPLFGFAGASVB 530
QY 538 HKTQ 541
DB 531 AANQ 534

```

RESULT 10  
 AAB48439  
 ID AAB48439 standard; Protein; 532 AA.

XX AC ABA48439;  
 XX DT 05-FEB-2002 (first entry)  
 XX DE Listeria monocytogenes protein #1143.  
 XX KM Antibacterial; gene therapy; vaccine; biosynthesis; biodegradation;  
 XX KM vitamin B12; bacterial infection; disease.  
 XX OS Listeria monocytogenes.  
 XX PN MO200177335-A2.  
 XX PD 18-OCT-2001.  
 XX PF 11-APR-2001; 2001MO-FR01118.  
 XX PR 11-APR-2000; 2000FR-0004629.  
 XX PA (INSP) INST PASTER.  
 XX PI Buchrieser C, Frangoul L, Couve E, Rusniok C, Feihl H, Dehoux P,  
 XX PI Dussurge O, Chetouani F, Nedjari H, Glaser P, Kunst F, Cossart P,  
 XX PI Daniels J, Goebel W, Kreft J, Kuhn M, Ng E, Vazquez-Boland JA,  
 XX PI Dominguez-Bernal G, Garrido-Garcia P, Tierrez-Martinez A, Amend A,  
 XX PI Chakraborty T, Dommann E, Hain T, Berche P, Charbit A, Durant L,  
 XX PI Perez-Diaz J, Baquero F, Garcia Del Portillo F, Gomez-Lopez N,  
 XX PI Madueno E, De Pablos B, Weiland J, Kaerst U, Entian K, Hauf J,  
 XX PI Rose M, Voss H;  
 XX DR MPI; 2002-010914/01.  
 XX PT Genomic sequence for Listeria monocytogenes, useful e.g. for treatment  
 XX PT and prevention of Listeria and related bacterial infections, and  
 XX PT related polypeptides -  
 XX PS Claim 6; SEQ ID No 1144; 192pp; French.  
 XX SN The present invention relates to the genome sequence of Listeria  
 XX CC monocytogenes EGD-e (see ABA03041). The genome sequence and fragments of  
 XX CC it are useful for selecting probes and primers for detecting genes in L.  
 XX CC monocytogenes and related organisms, and for studying genetic  
 XX CC polymorphisms and other genomes. The present sequence is a protein  
 XX CC encoded by the genome sequence of the present invention. Proteins  
 XX CC expressed from the genome sequence are useful for raising specific  
 XX CC antibodies, identification of L. monocytogenes and related organisms, and  
 XX CC for biosynthesis and biodegradation, especially biosynthesis of Vitamin  
 XX CC B12. The genome sequence and proteins encoded by it are also useful for  
 XX CC selecting compounds that regulate gene expression and cell replication  
 XX CC and modulate L. monocytogenes-related diseases. In addition, the genome  
 XX CC sequence and proteins encoded by it are useful in pharmaceutical and  
 XX CC vaccine compositions for the treatment or prevention of infections by L.  
 XX CC monocytogenes and related organisms.  
 XX CC Note: The sequence data for this patent did not form part of the printed  
 XX CC specification, but was obtained in electronic format directly from WIPO  
 XX CC at ftp.wipo.int/pub/published\_pct\_sequences.  
 XX SQ Sequence 532 AA;

Query Match 57.4%; Score 1600.5; DB 23; Length 532;  
 Best Local Similarity 55.2%; Pred. No. 1.2e-138;  
 Matches 299; Conservative 104; Mismatches 126; Indels 13; Gaps 3;

QY 1 MTKFPIPTGGVSSLGKGIASGLAALERGLKVTITKLPYINVDPRGMSFPGGEV 60  
 DB 1 MTKYIFVTGGVSSIGGITAASIGRLKNRGLSVITQKDPYINVDPRGMSFYQGEV 60  
 QY 61 VTEDGATDDLDGHERFLKTKTKKNNFTTGGVREOVLENKRGYLGATVVIPIHND 120  
 DB 61 VTDDGATDDLDGHERFLKTKTKKNNFTTGGVREOVLENKRGYLGATVVIPIHND 120  
 QY 121 EIKRRVYESAE--GKVALIEVGSTVGDISLSPLEETIRQMGVELGRDRALFTHLTVPY 178

DB 121 ELKORVFAAARMTNSDIITTEIGTVDIESLPFEHAIRQIKDVGAAENVLYIHITLIPY 180  
 QY 179 IKSAGELTKPTQSHSKELRTTGTIOPTILCRSPQIPASERRKIALFTNVAEKAVISAI 238  
 DB 181 IKAAGEKTKPTQSHSKELRTTGTIOPTILCRSPQIPASERRKIALFTNVAEKAVISAI 240  
 QY 239 DADTYRIPILRREGGLDVLVDOLRPPAADLSAMEKVVDGLPHPTDEYSIAIVGKVY 298  
 DB 241 DEETLYNPPLSLQKMDIVLEHLQEAPOEMTDWKULVHRVKNLSKRAIGLVGKIV 300  
 QY 299 DHTDAVKSINELINAGIHTRHKKVQISYIDSETTEAEGTAR-LKNVDAILVPGGEGEV 357  
 DB 301 SLQDAVLSVAELRHAGYDHDABEIRIWDISEKVPVVAEIMKDVDCILVPGGEGEBAI 360  
 QY 358 EKKISTVFARENKTPYIGICGMSAITEFARNVVGEGASHTEFLKSPHPVIGLITE 417  
 DB 361 EKKIAITEFARNKVPYIGICGMSAITEFARNVVGEGASHTEFLKSPHPVIGLITE 420  
 QY 418 WMDEAGELVTRDESDLGTRMLGAQKRLKADSLAPOLYOKDVITERRHRHRENNQYL 477  
 DB 421 -----QKNIENMGTLRLGYPARIKQGTAAEAAGTTLVEERHRRHRENNQYR 470  
 QY 478 KOLEAAGKTFSGSKSIDGRUVBIIIEPBPFLACQFHPFTSTPKNHALLPSCGYEAANK 537  
 DB 471 EGMERAAKIVSATSPDGLVAVELIDHPFVACQYHPEFISRPNRPSLFDVGAALIK 530  
 QY 538 HK 539  
 DB 531 NK 532

RESULT 11  
 AAU36779  
 ID AAU36779 standard; Protein; 536 AA.

XX AC AAU36779;  
 XX DT 14-FEB-2002 (first entry)  
 XX DE Staphylococcus aureus cellular proliferation protein #949.  
 XX KM Antisense; prokaryotic cellular proliferation protein;  
 XX KM antibiotic; antibacterial; drug design.  
 XX OS Staphylococcus aureus.  
 XX PN WO200170955-A2.  
 XX PD 27-SEP-2001.  
 XX PF 21-MAR-2001; 2001MO-US09180.  
 XX PR 21-MAR-2000; 2000US-191078P.  
 XX PR 23-MAY-2000; 2000US-206848P.  
 XX PR 26-MAY-2000; 2000US-207727P.  
 XX PR 23-OCT-2000; 2000US-242578P.  
 XX PR 27-NOV-2000; 2000US-253625P.  
 XX PR 22-DEC-2000; 2000US-257931P.  
 XX PR 16-FEB-2001; 2001US-269308P.  
 XX PA (ELIT-) ELITRA PHARM INC.  
 XX PI Haselbeck R, Ohlsen KL, Zyskind JW, Wall D, Trawick JD, Carr GT;  
 XX PI Yamamoto RT, Xu HH;  
 XX DR MPI; 2001-611495/70.  
 XX DR N-PSDB; AAS54638.  
 XX PT New polymucleotides for the identification and development of  
 XX PT antibiotics, comprise sequences of antisense nucleic acids -  
 XX PS Example 3; Seq ID No 12372; 511pp; English.

XX The invention relates to antisense inhibitors of genes essential to  
 CC prokaryotic cellular proliferation, their use in identifying the  
 CC genes, their use in the discovery of novel antibiotics, the essential  
 CC genes themselves and the encoded proteins. The prokaryotes used are  
 CC *Escherichia coli*, *Staphylococcus aureus*, *Salmonella typhi*, *Klebsiella*  
 CC *pneumoniae*, *Pseudomonas aeruginosa* and *Enterococcus faecalis*. The  
 CC invention is also useful for the identification of potential new targets  
 CC for antibiotic development. The antisense nucleic acids can also be used  
 CC to identify proteins used in proliferation, to express these proteins,  
 CC and to obtain antibodies capable of binding to the expressed proteins.  
 CC The proteins can be used to screen compounds in rational drug discovery  
 CC programmes. The antisense nucleic acid sequence is also useful to screen  
 CC for homologous nucleic acids which are required for cell proliferation in  
 CC a wide variety of organisms. The present sequence represents an  
 CC essential prokaryotic cellular proliferation protein.  
 CC Note: The sequence data for this patent did not form part  
 CC of the printed specification, but was obtained in electronic  
 CC format directly from WIPO at  
 CC ftp.wipo.int/pub/published\_pct\_sequences.

XX Sequence 536 AA;

Query Match 56.5%; Score 1576; DB 22; Length 536;

Best Local Similarity 55.6%; Pred. No. 2.2e-136;

Matches 302; Conservative 93; Mismatches 134; Indels 14; Gaps 4;

```

1 MTKFIFITGGVSSLSGKGIASSLAILEDGKLTITLDPYINVDPTGMSPPQHGVEF 60
1 MTKFIFITGGVSSLSGKGIASSLAILEDGKLTITLDPYINVDPTGMSPPQHGVEF 60
61 VTEDGAEITDLDGHERFLKTTMTKKNFTTGOVVEQVLRNKRKDYIGATVQVPHITD 120
61 VTDDGAEITDLDGHERFLIDINLNKFSNTAGKVSHTLKKERBGDYIGATVQVPHITN 120
121 EIKRRVYESAE--GKDVALLVGGTGGVGDIESLPLETTIKMGVEIGDRALPHILTVPY 178
121 EIKERLLLGESTNADVVITTEGGTGGDIESLPFEIARQISDLDGRNVMVHCTLLPY 180
179 IKSAGELTKPTQHSVKELENTIGIQPDILICSEQPIPAESERRKIALFTNVAEKAVISAI 238
181 IKSAGEMTKRTPQHSVKELENTIGIQPDILIVTEYMTQDMDKIALPCDINKESVIBOR 240
239 DADTIRIPLLRBQGLDVLVDQRLDVP--AADLSAMKRVVDGLTHPTDVSIAIVKRY 297
241 DADSLYEIPIQLSQQNMDDIVIKRQILNKNKYETQDLEMKQLDIVNNLDGKTTIGVGY 300
298 VDHETDVKSLNREALIHAGIHTHRKVOISYIDSEFTIEABGTAK--LKNVDAILVPGGFGERG 356
301 VSLQDVAIVSVESLKRKGYPPAKQDIDIRNIDSEVTDENAAETLADVDGILVPGGFGFRA 360
357 VEGKISTVFARENKIPYLGICLGMQSAVIEFARNVNGEGASHTFELPKSPHVIIGLT 416
361 SEGRISAIRKARENNVPPFGICLGMQLATVEFRNVLGEGASAEALDPATPYITDILP 420
417 EMMDZAGELVTRNEDDGLGTMRIGOKRKLKADSLAPQLKQDVITERRHRYERNQY 476
421 EQDQ-----IDLGGLTLRLGLYPCSIKRGTLAQGVGAALTEHRHRYERNNDY 470
477 LKOLEAAGKMGFSKSLDGRLEVEIIELEPHMFLACFHPSEFTSPENGALFSGVPEAA 536
471 RBLQLEANGVITSTSPDGRLEVENVEIPTNDFLACQFHHBFLSRPHRPPIKSFIMBL 530
471 RBLQLEANGVITSTSPDGRLEVENVEIPTNDFLACQFHHBFLSRPHRPPIKSFIMBL 530
537 KHK 539
531 KYQ 533

```

RESULT 12  
 AAU37385  
 AAU37385 standard; Protein; 536 AA.  
 XX  
 AC AAU37385;

XX 14-FEB-2002 (first entry)  
 DT Staphylococcus aureus cellular proliferation protein #1555.  
 DE  
 XX Antisense; prokaryotic cellular proliferation protein;  
 KW antibiotic; antibacterial; drug design.  
 XX  
 OS Staphylococcus aureus.  
 XX  
 PN WO200170955-A2.  
 XX  
 PD 27-SEP-2003.  
 XX  
 PF 21-MAR-2001; 2001WO-US09180.  
 XX  
 PR 21-MAR-2000; 2000US-191078P.  
 PR 23-MAY-2000; 2000US-206488P.  
 PR 26-MAY-2000; 2000US-207272P.  
 PR 23-OCT-2000; 2000US-242578P.  
 PR 27-NOV-2000; 2000US-253625P.  
 PR 22-DEC-2000; 2000US-257931P.  
 PR 16-FEB-2001; 2001US-269308P.  
 XX  
 PA (ELIT-) ELITRA PHARM INC.  
 XX  
 PI Haselbeck R, Ohlsen KU, Zyskind JW, Wall D, Trawick JD, Carr GJ;  
 PI Yamamoto RT, Xu HH;  
 XX  
 DR WPI, 2001-611495/70.  
 DR N-PSDB; AAS55244.  
 XX  
 PT New polynucleotides for the identification and development of  
 PT antibiotics, comprise sequences of antisense nucleic acids -  
 XX  
 RS Example 3; Seq ID No 12978; 511pp; English.  
 XX  
 CC The invention relates to antisense inhibitors of genes essential to  
 CC prokaryotic cellular proliferation, their use in identifying the  
 CC genes, their use in the discovery of novel antibiotics, the essential  
 CC genes themselves and the encoded proteins. The prokaryotes used are  
 CC *Escherichia coli*, *Staphylococcus aureus*, *Salmonella typhi*, *Klebsiella*  
 CC *pneumoniae*, *Pseudomonas aeruginosa* and *Enterococcus faecalis*. The  
 CC invention is also useful for the identification of potential new targets  
 CC for antibiotic development. The antisense nucleic acids can also be used  
 CC to identify proteins used in proliferation, to express these proteins,  
 CC and to obtain antibodies capable of binding to the expressed proteins.  
 CC The proteins can be used to screen compounds in rational drug discovery  
 CC programmes. The antisense nucleic acid sequence is also useful to screen  
 CC for homologous nucleic acids which are required for cell proliferation in  
 CC a wide variety of organisms. The present sequence represents an  
 CC essential prokaryotic cellular proliferation protein.  
 CC Note: The sequence data for this patent did not form part  
 CC of the printed specification, but was obtained in electronic  
 CC format directly from WIPO at  
 CC ftp.wipo.int/pub/published\_pct\_sequences.

XX Sequence 536 AA;

Query Match 56.5%; Score 1576; DB 22; Length 536;

Best Local Similarity 55.6%; Pred. No. 2.2e-136;

Matches 302; Conservative 93; Mismatches 134; Indels 14; Gaps 4;

```

1 MTKFIFITGGVSSLSGKGIASSLAILEDGKLTITLDPYINVDPTGMSPPQHGVEF 60
1 MTKFIFITGGVSSLSGKGIASSLAILEDGKLTITLDPYINVDPTGMSPPQHGVEF 60
61 VTEDGAEITDLDGHERFLKTTMTKKNFTTGOVVEQVLRNKRKDYIGATVQVPHITD 120
61 VTDDGAEITDLDGHERFLIDINLNKFSNTAGKVSHTLKKERBGDYIGATVQVPHITN 120
121 EIKRRVYESAE--GKDVALLVGGTGGVGDIESLPLETTIKMGVEIGDRALPHILTVPY 178

```

Db 121 EIKERLLAGESTNADVAVTEIGTGTGDIISLPPIAIRQISDGRBNVYVHCTLLPY 180  
 QY 179 IKSAGELTKPTQHSVKELRTIGIOPDILICRSEQPIPASERKIALFTNVAEKATISAI 238  
 Db 181 IKAAGEMTKPTQHSVKELRTIGIOPDILIVRTEYEMTODLKOKIALFODINKSEVIECR 240  
 QY 239 DADTIYRPLRLRBOGLDVLVDQLRLDVP--AADLSAMEKXVGLTHPTDEVSAIATVGYK 297  
 Db 241 DADSLYEIPQLISQNMDDIVIKRLOLNAYETQLDEWQQLDVIANNLQKITIGLVGKY 300  
 QY 298 VDHDTAYKSLNEMLIHAGITFRKVOISYIDSETIEAGTAK--LKNVDAILVPGGGERG 356  
 Db 301 VSLQDAVLSVSESLKAGYFPARDIDIRWIDSEBVDENAAEYLAQVDGILVPGGGERG 360  
 QY 357 VEGKISTVRPARKNKIPYLGICLQMOGSAVIEPARNVGLGAGSTFELPKSPHVPGLIT 416  
 Db 361 SEBKISALIKARKNNVFPFGICLQMOGLATVEFSRNVLGEGHNSALDPAITPPIIDLLP 420  
 QY 417 EMNDEAGELVTRDESDDLGSTMRLGAQCKRLKADSLAFLQYOKOVITERRHRYEFNNQY 476  
 Db 421 EOKD-----IEDLGITRLGLYPCSIKEGTILAQVYGKALIEERRHRYEFNNQY 470  
 QY 477 LKQLEAAGMKFSKSLDGRLEVEIIELEPHFWPLACQHPPEFTSPENGHALPFGFVAAA 536  
 Db 471 REQLESNGAVSGLSPDGRLEVEIIEIPNDFFIACQFHPPEFTSPRPHPIFKSFVEAL 530  
 QY 537 KHK 539  
 Db 531 NYQ 533

# RESULT 13

AA881965 standard; Protein: 535 AA.

AC AA881965;  
 AC XX  
 DT 03-SEP-2001 (first entry)  
 XX  
 DE S. epidermidis open reading frame protein sequence SEQ ID NO:1024.  
 XX  
 KM Staphylococcus epidermidis SRI strain; infection; diagnosis;  
 KM vaccination; endocarditis.  
 XX  
 OS Staphylococcus epidermidis.  
 XX  
 PN MO200134809-A2.  
 XX  
 PD 17-MAY-2001.  
 PF 09-NOV-2000; 2000MC-US30782.  
 PR 09-NOV-1999; 99US-0164258.  
 PA (GLAX ) GLAXO GROUP LTD.  
 XX  
 PI Kimmery M.;  
 XX  
 XX WPI, 2001-316495/33.  
 DR N-PSDB; AAH52815.  
 XX  
 PT Nucleic acids encoding polypeptides from Staphylococcus epidermidis,  
 PT useful for vaccinating against infections, e.g. endocarditis -  
 XX  
 PS Claim 18; Page 297-298; 2188pp; English.  
 CC AAH52304 to AAH53970 represent nucleic acids (I) encoding polypeptides  
 CC (II), given in AA881965 to AA881970, from Staphylococcus epidermidis.  
 CC (I) and (II) can have antibacterial activity and therefore can be used  
 CC in vaccination. The nucleic acids (I) may be used to produce the  
 CC S. epidermidis polypeptides (II) via the production of vectors  
 CC containing them which are used to produce host cells which express the  
 CC polypeptides. The polypeptides (II) (and/or nucleic acids) may then be

CC used to vaccinate subjects and to raise antibodies against the bacteria.  
 CC The polypeptides may also be used to assay for other inhibitors of their  
 CC activity and therefore identify compounds that may be used for the  
 CC treatment of S. epidermidis infections, e.g. endocarditis. AAH53971 to  
 CC AAH55090 represent specifically claimed S. epidermidis genomic DNA  
 CC polynucleotide sequences from the present invention. AAH55091 to  
 CC AAH55098 represent oligonucleotide sequences and primers which are used  
 CC in the exemplification of the present invention.  
 CC N.B. The present invention specifically claims all the polynucleotide  
 CC sequences given in the sequence listing of the present specification,  
 CC however the sequence listing only goes up to SEQ ID NO:4454 so even  
 CC though sequences are given in the disclosure for SEQ ID NO:4465 to 4472,  
 CC no sequences are present for SEQ ID NO:4455 to 4464.

Sequence 535 AA;

Query Match 56.5%; Score 1575; DB 22; Length 535;  
 Best Local Similarity 55.6%; Pred. No. 2.7e-136;  
 Matches 302; Conservative 95; Mismatches 132; Indels 14; Gaps 4;

QY 1 MTKFFITGVVSSLGKGIASSLAAILDEKGLKVTIKLPYINVDGMSPFQGHVF 60  
 Db 1 MTKFFITGVVSSLGKGIASSLAAILDEKGLKVTIKLPYINVDGMSPFQGHVF 60  
 QY 61 VTEDGAEITDLDLGHYERFLKTTMTKNNFTTGQYVEQVLAERKGDYLGATVQVPIHITD 120  
 Db 61 VTEDGAEITDLDLGHYERFLKTTMTKNNFTTGQYVEQVLAERKGDYLGATVQVPIHITD 120  
 QY 121 EIKRRVSEAE--GNDVALIEVGTVGDIISLPFETIRQMGVELGRALFIHULTVY 178  
 Db 121 EIKERLLAGESTNADVAVTEIGTGTGDIISLPPIAIRQISDGRBNVYVHCTLLPY 180  
 QY 179 IKSAGELTKPTQHSVKELRTIGIOPDILICRSEQPIPASERKIALFTNVAEKATISAI 238  
 Db 181 IKAAGEMTKPTQHSVKELRTIGIOPDILIVRTEYEMTODLKOKIALFODINKSEVIECR 240  
 QY 239 DADTIYRPLRLRBOGLDVLVDQLRLDVP--AADLSAMEKXVGLTHPTDEVSAIATVGYK 297  
 Db 241 DADSLYEIPQLISQNMDDIVIKRLOLNAYETQLDEWQQLDVIANNLQKITIGLVGKY 300  
 QY 298 VDHDTAYKSLNEMLIHAGITFRKVOISYIDSETIEAGT--AKLNVDAILVPGGGERG 356  
 Db 301 VSLQDAVLSVSESLKAGYFPARDIDIRWIDSEBVDENAAEYLAQVDGILVPGGGERG 360  
 QY 357 VEGKISTVRPARKNKIPYLGICLQMOGSAVIEPARNVGLGAGSTFELPKSPHVPGLIT 416  
 Db 361 SEBKISALIKARKNNVFPFGICLQMOGLATVEFSRNVLGEGHNSALDPAITPPIIDLLP 420  
 QY 417 EMNDEAGELVTRDESDDLGSTMRLGAQCKRLKADSLAFLQYOKOVITERRHRYEFNNQY 476  
 Db 421 EOKD-----IEDLGITRLGLYPCSIKEGTILAQVYGKALIEERRHRYEFNNQY 470  
 QY 477 LKQLEAAGMKFSKSLDGRLEVEIIELEPHFWPLACQHPPEFTSPENGHALPFGFVAAA 536  
 Db 471 REQLESNGAVSGLSPDGRLEVEIIEIPNDFFIACQFHPPEFTSPRPHPIFKSFVEAL 530  
 QY 537 KHK 539  
 Db 531 NYQ 533

# RESULT 14

ABP39346 standard; Protein: 561 AA.

AC ABP39346;  
 AC XX  
 DT 24-JUL-2002 (first entry)  
 XX  
 DE Staphylococcus epidermidis ORF amino acid sequence SEQ ID NO:4191.  
 XX  
 KM Staphylococcus epidermidis, open reading frame; ORF; bacterial infection;  
 KM antibacterial; gene therapy.



Query Match 56.3%; Score 1569; DB 22; Length 536;  
Best local Similarity 55.4%; Pred. No. 9,7e-136;  
Matches 301; Conservative 94; Mismatches 134; Indels 14; Gaps 4;

```
QY 1 MTKYIFVTGGVSSLGKGIASSLAAILLEDRLKVTITKLDPIYINVDGTMSPPOHGEVF 60
Db 1 MTKYIFVTGGVSSLGKGIASSLAAILLEDRLKVTITKLDPIYINVDGTMSPPOHGEVF 60
QY 61 VTEDGATDLDLIGYERFLKTMKKNFTTGOVYBOYLNERKGDYLGATVQVPIPHITD 120
Db 61 VTEDGATDLDLIGYERFLKTMKKNFTTGOVYBOYLNERKGDYLGATVQVPIPHITD 120
QY 121 EIKRVTESAE--GKDYALIEVGGTVGDIESLPLEFTIRQMGVELGRDALFTHLTVPY 178
Db 121 EIKRVTESAE--GKDYALIEVGGTVGDIESLPLEFTIRQMGVELGRDALFTHLTVPY 178
QY 121 EIKRVTESAE--GKDYALIEVGGTVGDIESLPLEFTIRQMGVELGRDALFTHLTVPY 180
Db 121 EIKRVTESAE--GKDYALIEVGGTVGDIESLPLEFTIRQMGVELGRDALFTHLTVPY 180
QY 179 IKSNGELKTPQHSVVELTGTIGQPDILICRSGOPTPASERKIALFTVAEKAVISAI 238
Db 179 IKSNGELKTPQHSVVELTGTIGQPDILICRSGOPTPASERKIALFTVAEKAVISAI 238
QY 181 LKAGNRKTPQHSVVELTGTIGQPDILICRSGOPTPASERKIALFTVAEKAVISAI 240
Db 181 LKAGNRKTPQHSVVELTGTIGQPDILICRSGOPTPASERKIALFTVAEKAVISAI 240
QY 239 DADTYRIPLLEBOGLDLVDQLRDVPADLSAWEKVVVDGLTPDEVSAIVGKXY 298
Db 239 DADTYRIPLLEBOGLDLVDQLRDVPADLSAWEKVVVDGLTPDEVSAIVGKXY 298
QY 241 DVEETVSIPLALQONNDQIVCHDKIDAPADMTERALAEKYNLKKTKRIALVGYV 300
Db 241 DVEETVSIPLALQONNDQIVCHDKIDAPADMTERALAEKYNLKKTKRIALVGYV 300
QY 299 DHTDAYKSLNBAIHHAGIHTRHKVQISYIDSETTEABGTA-KLKNVDAILVPGGFERGV 357
Db 299 DHTDAYKSLNBAIHHAGIHTRHKVQISYIDSETTEABGTA-KLKNVDAILVPGGFERGV 357
QY 301 ELPDAYSIVVEALKHAGFDPDSIDIEIDWVDSQELTAEVNERIGSADGILVPGGFERGI 360
Db 301 ELPDAYSIVVEALKHAGFDPDSIDIEIDWVDSQELTAEVNERIGSADGILVPGGFERGI 360
QY 358 ESKISTVFARBNKIPYLGICLGMQSAVIEFARNVVGLEGAHSTEFLEKSPHVLGLITE 417
Db 358 ESKISTVFARBNKIPYLGICLGMQSAVIEFARNVVGLEGAHSTEFLEKSPHVLGLITE 417
QY 361 ESKIEAIRFARNVDPPLGICLGMQSAVIEFARNVVGLEGAHSTEFLEKSPHVLGLITE 420
Db 361 ESKIEAIRFARNVDPPLGICLGMQSAVIEFARNVVGLEGAHSTEFLEKSPHVLGLITE 420
QY 418 WMDAGELVTRDSDSDIGSTRLAGOKRLKADSLAFOLY-QKQVITERRHRYEENNOY 476
Db 418 WMDAGELVTRDSDSDIGSTRLAGOKRLKADSLAFOLY-QKQVITERRHRYEENNOY 476
QY 421 -----QENIENLGGTLRLGLVPCCLKKGTAAAGNEDVQERHRYEENNOY 470
Db 421 -----QENIENLGGTLRLGLVPCCLKKGTAAAGNEDVQERHRYEENNOY 470
QY 477 LKQLEAAGKMFSGKSLDGLVETIELEPHMPLACQHPHPTSTPRNGHALFSGFEEA 536
Db 477 LKQLEAAGKMFSGKSLDGLVETIELEPHMPLACQHPHPTSTPRNGHALFSGFEEA 536
QY 471 RQLEENGVLVSGVSPDNLVEIVEIPEKQFVACQFHPHLSRPNRPORLKGFVGAAL 530
Db 471 RQLEENGVLVSGVSPDNLVEIVEIPEKQFVACQFHPHLSRPNRPORLKGFVGAAL 530
QY 537 KHK 539
Db 531 ANK 533
```

Search completed: January 29, 2004, 15:49:45  
Job time : 51.7262 secs

GenCore version 5.1.6  
Copyright (c) 1993 - 2004 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: January 29, 2004, 15:45:34 ; Search time 17.0734 Seconds  
(without alignments)  
1346.130 Million cell updates/sec

Title: US-09-941-947a-16

Perfect score: 2788

Sequence: 1 MKTFIPITGVSSLSKQIA.....HALFSGFEVAAKHTQGTG 544

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 328717 seqs, 42310858 residues

Total number of hits satisfying chosen parameters: 328717

Minimum DB seq length: 0  
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%  
Listing first 45 summaries

Database :

1: Issued Patents AA:\*  
2: /cgn2\_6/prodata/1/iaa/5A\_COMB.pep:\*  
3: /cgn2\_6/prodata/1/iaa/5B\_COMB.pep:\*  
4: /cgn2\_6/prodata/1/iaa/6A\_COMB.pep:\*  
5: /cgn2\_6/prodata/1/iaa/6B\_COMB.pep:\*  
6: /cgn2\_6/prodata/1/iaa/6C\_COMB.pep:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

#### SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	2004	71.9	547	4	US-09-252-991A-21120
2	1825.5	65.5	567	4	US-09-328-352-7224
3	1575	56.5	561	4	US-09-134-001C-4191
4	1355.5	48.6	537	4	US-09-198-452A-246
5	1038	37.3	340	4	US-09-107-532A-6925
6	544.5	19.5	226	4	US-09-107-532A-4108
7	220.5	7.9	373	4	US-09-252-991A-25369
8	123.5	4.4	830	4	US-09-107-532A-5709
9	120	4.3	401	4	US-09-252-991A-27277
10	111.5	4.0	552	3	US-08-116-098-2
11	111.5	4.0	552	3	US-08-687-590-32
12	111.5	4.0	619	3	US-09-066-046-2
13	109.5	3.9	1407	4	US-09-328-352-7885
14	109.5	3.9	809	4	US-09-252-991A-3742
15	107.5	3.9	1700	4	US-09-252-991A-21763
16	107.5	3.8	773	4	US-09-252-991A-31403
17	106	3.8	171	2	US-08-560-098A-53
18	105.5	3.8	887	4	US-09-252-991A-16679
19	105	3.8	175	4	US-09-134-001C-4747
20	103.5	3.7	573	4	US-09-328-352-6420
21	103	3.7	483	4	US-08-887-534A-51
22	103	3.7	483	4	US-09-527-431-51
23	103	3.7	804	4	US-09-134-001C-5218
24	102.5	3.7	644	4	US-09-252-991A-25084
25	102.5	3.7	683	4	US-09-252-991A-32144
26	102	3.7	366	4	US-09-134-001C-4799
27	102	3.7	783	6	5231168-2

28	102	3.7	1027	3	US-08-446-137B-2	Sequence 2, Appl
29	101.5	3.6	458	4	US-09-198-452A-610	Sequence 610, App
30	101	3.6	236	3	US-08-935-263-2	Sequence 2, Appl
31	101	3.6	236	4	US-09-594-185-2	Sequence 2, Appl
32	101	3.6	775	1	US-07-603-133B-13	Sequence 13, Appl
33	100.5	3.6	582	4	US-09-252-991A-26782	Sequence 26782, A
34	100.5	3.6	1151	3	US-08-840-006-6	Sequence 6, Appl
35	100.5	3.6	1200	3	US-08-840-006-5	Sequence 5, Appl
36	100.5	3.6	3724	2	US-08-804-227C-10	Sequence 10, Appl
37	100.5	3.6	3724	2	US-08-804-198-4	Sequence 4, Appl
38	100	3.6	1044	4	US-09-252-991A-24495	Sequence 24495, A
39	100	3.6	1802	4	US-09-322-478-18	Sequence 18, Appl
40	99.5	3.6	524	4	US-09-252-991A-27783	Sequence 27783, A
41	99.5	3.6	551	4	US-09-252-991A-24209	Sequence 24209, A
42	99.5	3.6	858	4	US-09-255-828-22	Sequence 22, Appl
43	99.5	3.6	858	4	US-09-255-828-29	Sequence 29, Appl
44	99.5	3.6	921	4	US-09-252-991A-32974	Sequence 32974, A
45	99.5	3.6	1169	4	US-09-255-828-20	Sequence 20, Appl

#### ALIGNMENTS

RESULT 1  
US-09-252-991A-21120  
; Sequence 21120, Application US/09252991A  
; Patent No. 6551795  
; GENERAL INFORMATION:  
; APPLICANT: Marc J. Rubenfield et al.  
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO PSEUDOMONAS  
; TITLE OF INVENTION: AERUINOSA FOR DIAGNOSTICS AND THERAPEUTICS  
; FILE REFERENCE: 107196.136  
; CURRENT APPLICATION NUMBER: US/09/252,991A  
; PRIOR FILING DATE: 1999-02-18  
; PRIOR APPLICATION NUMBER: US 60/074,788  
; PRIOR FILING DATE: 1998-02-18  
; PRIOR APPLICATION NUMBER: US 60/094,190  
; PRIOR FILING DATE: 1998-07-27  
; NUMBER OF SEQ ID NOS: 33142  
; SEQ ID NO 21120  
; LENGTH: 547  
; TYPE: PRT  
; ORGANISM: Pseudomonas aeruginosa  
US-09-252-991A-21120

Query Match	71.9%	Score 2004;	DB 4;	Length 547;
Best Local Similarity	71.2%	Pred. No. 2,4e-188;		
Matches 383;	Conservative 63;	Mismatches 92;	Indels 0;	Gaps 0;
QY	1	MTKFIPTGVSSLSKQIAASSIAAIIEDGKXKTKLPYINVDGTPSPGHGEVF	60	
DB	6	MTRYIFVGVSSLSKQIASIASIAAIIEDGKXKTKLPYINVDGTPSPGHGEVF	65	
QY	61	VTEDAEVDLDGHERFLKTTMKNNFTTGOVYEVYLNERKGDYVATQVPHITD	120	
DB	66	VTDGAETDLDGHERFLKTTMKNNFTTGOVYEVYLNERKGDYVATQVPHITD	125	
QY	121	EIKRVRVSAAGKVALIEVGCTVGDHSPLFTIRGMVGLGRALPHLITVPIIK	180	
DB	126	EIKRRIIAGAGADVALIEIGTVGDHSPLFTIRGMVGLGRALPHLITVPIIA	185	
QY	181	SAGELKTPGTHSVETRTIGIOPDILCRSROPAPSERKIALFTNVAEASVSAIDA	240	
DB	186	TAGEKTFPGTHSVETRTIGIOPDILCRSROPAPSERKIALFTNVAEASVSAIDA	245	
QY	241	DTIYRIPLLREOGIDLDVVDLRVADPAIDAMERKVDGLHPTDEVSIAIVKRYVDH	300	
DB	246	DTIYRIPVLAAGDIDVVRPGLGCGADISEMDRVADKLPREVEVITAMVCKYVEL	305	
QY	301	TDAYKSLNEALIHGCHTRHVOVSYSIDSTFTEAGTKAKONVALIIVPGFGSGVGEK	360	
DB	306	LDAYKSLNEALIHGCHTRHVOVSYSIDSTFTEAGTKAKONVALIIVPGFGSGVGEK	365	

QY 361 ISTVFARENKIPYLGICLQMSQAVIFBARNVVLGAHSTELPKSPHPIVIGLITWMD 420  
 DB 366 ISTVFARENKIPYLGICLQMSQAVIFBARNVVLGAHSTELPKSPHPIVIGLITWMD 425  
 QY 421 EAGELVTRDSDSLGGMTRGLGAQCKRLKADSLAQLYQKQVITERHRRHREFNNQYLKOL 480  
 DB 426 ATGATBTRTESDSJGGMTRGLGAQCKRLKADSLAQLYQKQVITERHRRHREFNNQYLKOL 485  
 QY 481 EAAQMKESGKSLDRLVLEIELPEHPWFLACQFPEPTSTPRNGHALFSGFVEAANA 538  
 DB 486 EQAGLXISGSGDGLAVVEVSEPHFWACQFPEPTSTPRNGHALFSGFVEAANA 543

# RESULT 2 US-09-328-352-7224

; Sequence 7224, Application US/09328352  
 ; Patent No. 6562958  
 ; GENERAL INFORMATION:  
 ; APPLICANT: Gary L. Bretton et al.  
 ; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO ACINETOBACTER  
 ; FILE REFERENCE: GTC99-03PA  
 ; CURRENT APPLICATION NUMBER: US/09/328,352  
 ; CURRENT FILING DATE: 1999-06-04  
 ; NUMBER OF SEQ ID NOS: 8252  
 ; SEQ ID NO 7224  
 ; LENGTH: 567  
 ; TYPE: PRN  
 ; ORGANISM: Acinetobacter baumannii  
 ; US-09-328-352-7224

Query Match 65.5%; Score 1825.5; DB 4; Length 567;  
 Best Local Similarity 64.6%; Pred. No. 9,3e-171;  
 Matches 351; Conservative 81; Mismatches 106; Indels 5; Gaps 3;

QY 1 MTKFPIFGVSSVSGKGIASLAAIILEDRLKKTITKDPYINVDPGTMSFPGHGF 60  
 DB 23 MTKFPIFGVSSVSGKGIASLAAIILEDRLKKTITKDPYINVDPGTMSFPGHGF 82  
 QY 61 VTEDGATDLDLGHYERFL-KTMTKKNFTTQVVEQVLRNKKDYLGAIVQVPHIT 119  
 DB 83 VTEDGATDLDLGHYERFL-KTMTKKNFTTQVVEQVLRNKKDYLGAIVQVPHIT 142  
 QY 120 DEIKRKYESAEGDVALIEVGGTVGDIESLPLETRKMGVLEGDRALFIHLTLVPI 179  
 DB 143 DEIKRKYESAEGDVALIEVGGTVGDIESLPLETRKMGVLEGDRALFIHLTLVPI 202  
 QY 180 KSAGELKTKPTQSHVKEKELRTIGIOPDILICRSEOPIPASERKIKALFTVAERKVAISAD 239  
 DB 203 KSAGELKTKPTQSHVKEKELRTIGIOPDILICRSEOPIPASERKIKALFTVAERKVAISAD 262  
 QY 240 ADITTRPLILREOGDLDLVVDQLR-DVPAADLSAMEKRVYDGLTHPTDEVSIAIVGKY 298  
 DB 263 AKTIYQIPRGFTYQNVDDLICERFGFTDLEADLTMDNVVALLPEYTVRAVMGKTV 322  
 QY 299 DHTDAYKSLNEALIHAGIHTRHVKQISYIDSETEAEGTAKLKNVDAILVPGFGRGYE 358  
 DB 323 ELDPAYSVNEALIHAGIHTRHVKQISYIDSETEAEGTAKLKNVDAILVPGFGRGYE 382  
 QY 359 GKSTVFARENKIPYLGICLQMSQAVIFBARNVVLGAHSTELPKSPHPIVIGLITW 418  
 DB 383 GKSTVFARENKIPYLGICLQMSQAVIFBARNVVLGAHSTELPKSPHPIVIGLITW 442  
 QY 419 MDAGELVTRDSDSLGGMTRGLGAQCKRLKADSLAQLYQKQVITERHRRHREFNNQYLK 478  
 DB 443 LDRGELVTRDSDSLGGMTRGLGAQCKRLKADSLAQLYQKQVITERHRRHREFNNQYLK 502  
 QY 479 QLEAAGKSGKSLDRLVLEIELPEHPWFLACQFPEPTSTPRNGHALFSGFVEAANA 538  
 DB 503 QLEAAGKSGKSLDRLVLEIELPEHPWFLACQFPEPTSTPRNGHALFSGFVEAANA 559  
 QY 539 KTG 541

DB 560 KTG 562

# RESULT 3 US-09-134-001C-4191

; Sequence 4191, Application US/09134001C  
 ; Patent No. 6380370  
 ; GENERAL INFORMATION:  
 ; APPLICANT: Lynn Doucette-Stamm et al.  
 ; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO STAPHYLOCOCCUS  
 ; FILE REFERENCE: GTC-007  
 ; CURRENT APPLICATION NUMBER: US/09/134,001C  
 ; PRIOR FILING DATE: 1998-08-13  
 ; PRIOR APPLICATION NUMBER: US 60/064,964  
 ; PRIOR FILING DATE: 1997-11-08  
 ; PRIOR APPLICATION NUMBER: US 60/055,779  
 ; NUMBER OF SEQ ID NOS: 5674  
 ; SEQ ID NO 4191  
 ; LENGTH: 561  
 ; TYPE: PRN  
 ; ORGANISM: Staphylococcus epidermidis  
 ; US-09-134-001C-4191

Query Match 56.5%; Score 1575; DB 4; Length 561;  
 Best Local Similarity 55.6%; Pred. No. 4e-146;  
 Matches 302; Conservative 95; Mismatches 132; Indels 14; Gaps 4;

QY 1 MTKFPIFGVSSVSGKGIASLAAIILEDRLKKTITKDPYINVDPGTMSFPGHGF 60  
 DB 27 MTKFPIFGVSSVSGKGIASLAAIILEDRLKKTITKDPYINVDPGTMSFPGHGF 86  
 QY 61 VTEDGATDLDLGHYERFL-KTMTKKNFTTQVVEQVLRNKKDYLGAIVQVPHIT 120  
 DB 87 VTEDGATDLDLGHYERFL-KTMTKKNFTTQVVEQVLRNKKDYLGAIVQVPHIT 146  
 QY 121 EIKRKYESAEGDVALIEVGGTVGDIESLPLETRKMGVLEGDRALFIHLTLVPI 178  
 DB 147 EIKRKYESAEGDVALIEVGGTVGDIESLPLETRKMGVLEGDRALFIHLTLVPI 206  
 QY 179 KSAGELKTKPTQSHVKEKELRTIGIOPDILICRSEOPIPASERKIKALFTVAERKVAISAD 238  
 DB 207 KSAGELKTKPTQSHVKEKELRTIGIOPDILICRSEOPIPASERKIKALFTVAERKVAISAD 266  
 QY 239 DADITTRPLILREOGDLDLVVDQLR-DVPAADLSAMEKRVYDGLTHPTDEVSIAIVGKY 297  
 DB 267 DADITTRPLILREOGDLDLVVDQLR-DVPAADLSAMEKRVYDGLTHPTDEVSIAIVGKY 326  
 QY 298 VHTDAYKSLNEALIHAGIHTRHVKQISYIDSETEAEGTAKLKNVDAILVPGFGRGYE 356  
 DB 327 VHTDAYKSLNEALIHAGIHTRHVKQISYIDSETEAEGTAKLKNVDAILVPGFGRGYE 386  
 QY 357 VEGKISTVFARENKIPYLGICLQMSQAVIFBARNVVLGAHSTELPKSPHPIVIGLIT 416  
 DB 387 VEGKISTVFARENKIPYLGICLQMSQAVIFBARNVVLGAHSTELPKSPHPIVIGLIT 446  
 QY 417 EMDAGELVTRDSDSLGGMTRGLGAQCKRLKADSLAQLYQKQVITERHRRHREFNNQYLK 476  
 DB 447 EMDAGELVTRDSDSLGGMTRGLGAQCKRLKADSLAQLYQKQVITERHRRHREFNNQYLK 502  
 QY 477 EMDAGELVTRDSDSLGGMTRGLGAQCKRLKADSLAQLYQKQVITERHRRHREFNNQYLK 536  
 DB 497 EMDAGELVTRDSDSLGGMTRGLGAQCKRLKADSLAQLYQKQVITERHRRHREFNNQYLK 556  
 QY 537 KTG 539  
 DB 557 KTG 559

RESULT 4  
 US-09-198-452A-246  
 ; Sequence 246, Application US/09198452A

Patent No. 6559294  
 GENERAL INFORMATION:  
 APPLICANT: Griffiths, R.  
 TITLE OF INVENTION: Chlamydia pneumoniae genomic sequence and polypeptides, fragments thereof and uses thereof, in particular for the diagnosis, prevention and treatment of infection  
 TITLE OF INVENTION: and treatment of infection  
 FILE REFERENCE: 9710-003-999  
 CURRENT APPLICATION NUMBER: US/09/198,452A  
 CURRENT FILING DATE: 1998-11-24  
 NUMBER OF SEQ ID NOS: 6849  
 SEQ ID NO 246  
 LENGTH: 537  
 TYPE: PRT  
 ORGANISM: Chlamydia pneumoniae  
 US-09-198-452A-246

Query Match 48.6%; Score 1355.5; DB 4; Length 537;  
 Best Local Similarity 50.5%; Pred. No. 1.5e-124;  
 Matches 271; Conservative 95; Mismatches 154; Indels 17; Gaps 7;

QY 3 KFIITGVVSSLGKIAASLAILEDGLKVTITKLDPIYINVDPTGMSPFQHGVEFVT 62  
 DB 4 KCIPIFGVYVSSLGKIAASLAILEDGLKVTITKLDPIYINVDPTGMSPFQHGVEFVT 63  
 QY 63 EDGAEITDLDGHERPLKTKMTKNNFTTGQYVEOYLAKERKDYIGATVOYIPHTTDEI 122  
 DB 64 DDGEITDLDGHERPLKTKMTKNNFTTGQYVEOYLAKERKDYIGATVOYIPHTTDEI 123  
 QY 123 KRAVESAE--GKVALLIEVGTVGDIESLPLETIRQMGVELGRDRALEFHLTVPYIK 180  
 DB 124 IQYILDAKESPDVILVEIGTIGIESLPLEIRQMGVELGRDRALEFHLTVPYIK 182  
 QY 181 SAGELTKPTQHSVKELEKTIQOPDILCRSEQPIPASRRKIALFTVAERKAVISADA 240  
 DB 183 AADVEKSKPTQHSVKELEKTIQOPDILCRSEQPIPASRRKIALFTVAERKAVISADA 242  
 QY 241 -DIIYIPLLRQGLDLDVVDQLRDVPAADLSAEKVVDDGTHETDEVSIAIVKXYD 299  
 DB 243 KHTIYEPMLAOEKLANPFGKELATVPENDRRVAVNQLSODLPYKIGVAGKYO 302  
 QY 300 HTDAYSLEALIHAGIHTRHAKVQISYIDSEITAEAGTAKLKNVDAILVPGRGGERVSG 359  
 DB 303 HRDAYSIFEALHTHALRLGHAAEILPIDAE--DENLIMELSGCQCLVPGRGGERVSG 360  
 QY 360 KISTVPAERKIPYIGICLQMSAVIEPAKRVGLEGHSTFLEKSPHYIGLITTEMK 419  
 DB 361 KIAAAKFCRQGIPIYIGICLQMSAVIEPAKRVGLEGHSTFLEKSPHYIGLITTEMK 416  
 QY 420 DEAGEIATDEDEDLDGTVELGAKCRKLRADSLAFOLY-OKOVITERHRRHYEFNNQYAK 478  
 DB 417 EGDDPLVA-----TGTRELGAIVPCLLKPSKAKHAYNESSLIQRHRRHYEFNNQYAK 470  
 QY 479 QLEAAGKTSKSLDLRLVETIELPEHPFPLACOPPEFTSTPRNGHALFSGFVEAA 535  
 DB 471 SLBDHGLRIATGTCPPQGLCEIIEVSDHPMWTIGVQFPRFVSXKLISSPHLFIATFMA 527

RESULT 5  
 US-09-107-532A-6925  
 Sequence 6925, Application US/09107532A  
 Patent No. 6583275

GENERAL INFORMATION:  
 APPLICANT: Lynn A Doucette-Stamm and David Bush  
 TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO ENTEROCOCCUS FAECIUM FOR DIAGNOSTICS AND THERAPEUTICS  
 NUMBER OF SEQUENCES: 7310  
 CORRESPONDENCE ADDRESS:  
 ADDRESSER: GENOME THERAPEUTICS CORPORATION  
 STREET: 100 Beaver Street  
 CITY: Waltham  
 STATE: Massachusetts  
 COUNTRY: USA  
 ZIP: 02354

COMPUTER READABLE FORM:  
 MEDIUM TYPE: CD/ROM ISO9660  
 COMPUTER: PC  
 OPERATING SYSTEM: <Unknown>  
 SOFTWARE: ASCII  
 CURRENT APPLICATION DATA:  
 APPLICATION NUMBER: US/09/107,532A  
 FILING DATE: 30-Jun-1998  
 PRIOR APPLICATION DATA:  
 APPLICATION NUMBER: 60/085,598  
 FILING DATE: 14 May 1998  
 APPLICATION NUMBER: 60/051571  
 FILING DATE: July 2, 1997  
 ATTORNEY/AGENT INFORMATION:  
 NAME: Arinello, Pamela Deneke  
 REGISTRATION NUMBER: 40,489  
 REFERENCE/DOCKET NUMBER: GTC-012  
 TELECOMMUNICATION INFORMATION:  
 TELEPHONE: (781)893-5007  
 TELEFAX: (781)893-8277  
 INFORMATION FOR SEQ ID NO: 6925:  
 SEQUENCE CHARACTERISTICS:  
 LENGTH: 340 amino acids  
 TYPE: amino acid  
 TOPOLOGY: linear  
 MOLECULE TYPE: protein  
 HYPOTHEICAL: YES  
 ORIGINAL SOURCE:  
 ORGANISM: Enterococcus faecium  
 FEATURE:  
 NAME/KEY: misc feature  
 LOCATION: (B) LOCATION 1...340  
 SEQUENCE DESCRIPTION: SEQ ID NO: 6925:  
 US-09-107-532A-6925

Query Match 37.3%; Score 1039; DB 4; Length 340;  
 Best Local Similarity 59.2%; Pred. No. 9.7e-94;  
 Matches 197; Conservative 61; Mismatches 63; Indels 12; Gaps 2;

QY 1 MTKFIITGVVSSLGKIAASLAILEDGLKVTITKLDPIYINVDPTGMSPFQHGVEF 60  
 DB 7 MTKYIFVGVVSSLGKIAASLAILEDGLKVTITKLDPIYINVDPTGMSPFQHGVEF 66  
 QY 61 VTDEGAEITDLDGHERPLKTKMTKNNFTTGQYVEOYLAKERKDYIGATVOYIPHTTDEI 120  
 DB 67 VTDEGAEITDLDGHERPLKTKMTKNNFTTGQYVEOYLAKERKDYIGATVOYIPHTTDEI 126  
 QY 121 EIKRIVESAE--GKVALLIEVGTVGDIESLPLETIRQMGVELGRDRALEFHLTVPY 178  
 DB 127 EIKERIKAAKWTQSDVILTEVGTVGDIESLPLETIRQMGVELGRDRALEFHLTVPY 186  
 QY 179 IKSAGELTKPTQHSVKELEKTIQOPDILCRSEQPIPASRRKIALFTVAERKAVISAI 238  
 DB 187 LKAAGEMTKPTQHSVKELEKTIQOPDILCRSEQPIPASRRKIALFTVAERKAVISAI 246  
 QY 239 DADITVIRIPLLRQGLDLDVVDQLRDVPAADLSAEKVVDDGTHETDEVSIAIVKXY 298  
 DB 247 DADITVIRIPLLRQGLDLDVVDQLRDVPAADLSAEKVVDDGTHETDEVSIAIVKXY 306  
 QY 299 DHTDAYSLEALIH-----AGIHTRHK 321  
 DB 307 ELDPAYISVEALKHLIVSPITQTSRISGRK 339

RESULT 6  
 US-09-107-532A-4108  
 Sequence 4108, Application US/09107532A  
 Patent No. 6583275

GENERAL INFORMATION:  
 APPLICANT: Lynn A Doucette-Stamm and David Bush  
 TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO ENTEROCOCCUS FAECIUM FOR DIAGNOSTICS AND THERAPEUTICS  
 NUMBER OF SEQUENCES: 7310



; INFORMATION FOR SEQ ID NO: 5709:  
 ; SEQUENCE CHARACTERISTICS:  
 ; LENGTH: 830 amino acids  
 ; TYPE: amino acid  
 ; TOPOLOGY: linear  
 ; MOLECULE TYPE: protein  
 ; HYPOTHEICAL: YES  
 ; ORIGINAL SOURCE:  
 ; ORGANISM: Enterococcus faecium  
 ; FEATURE:  
 ; NAME/KEY: misc feature  
 ; LOCATION: (8) LOCATION 1...830  
 ; SEQUENCE DESCRIPTION: SEQ ID NO: 5709:  
 US-09-107-532A-5709

Query Match 4.4%; Score 123.5; DB 4; Length 830;  
 Best Local Similarity 20.4%; Pred. No. 0.0051;  
 Matches 115; Conservative 80; Mismatches 201; Indels 169; Gaps 27;  
 15 LGKGIASSLAALIEDRGKLVITIKDPIYINVDPGTMSFONGEV--PYTE-----DG 65  
 258 MGLVAGTKYRGFEEDRMKV---IDBIYN-----DQVILFIDELHTLIGAG 303  
 66 AETDLDLGHYERFLKTTMKNNFTTGQV---YEOVLNKKRGDTLGTQVYIPIITDS 121  
 304 AEGALDASN---ILKDALAGELQITIGATTLDEYQYIKDSALERRFARICVDEPTPEE 360  
 122 IK-----RREYASAKGVALLIEGQVGDIESLPFLETIRMGVGLGRDLFIHTL 175  
 361 AEVILGLRKYREHHEGVELTDAVPAAYN--LSYNT--TSRL-----PDRAIDL-- 408  
 176 VPIKAGELKTKPTQHSVKELETTIGIQPDILICRSEQPIPASERRKIALFTNVV-- 230  
 409 --IDESAAKVRLLQTHLETKS--TVIKLEIDELVQEKKAIIQKOD-----PENAAQLRRQ 459  
 231 EKAV-----ISALDPTIRIPLLRBGLDVIYVDLADPADLSAMERK----- 278  
 460 EKALRKQLQKVSALBK-----QEQYSDRYER--DV--ATVSEMGVPLQOLE 506  
 279 -----VDLTHPTDEVSIAVGRYVDHTDAVYSINELIHAQIHRHKVQLSYIDS 329  
 507 KKESERLBELBGLHSE-----RVVQGEAVAVASRAIRRA----- 541  
 330 ETTIABGTAKLVND-----AIIVPGGEGRGVEBKISTVPRARENKIPIYIGICLAKMS 383  
 542 -----RSGIKNDPDRPIGSPFMTGPTGVKTEIATALEVMFGSDAL-----IRV 586  
 384 AVIEPARNVVGLGASTERFLPKSPHPIVGLITEMDBAGEIVTRDESDLGSTWELGAQ 443  
 587 DMESEFEK-----YSTSLISSPQYVG-----YDEGQLTEKIRKQPYSVIILDEVE 634  
 444 KCRILKADSLAFOLYQKDVITERHRRHYEFNN-----QYLKLEBAAGKFSGKSL 492  
 635 KAPDVFNILLQVLDGHLTDSGRKVDPRNTIMTMSIGATQIEEKVGVNVDVTK 694  
 493 DGLVLEILHPEHPWFLACQFHEP 517  
 695 DHRKQGRILE--EKKAFREP 715

RESULT 9  
 US-09-252-991A-27277  
 ; Sequence 27277 Application US/09252991A  
 ; Patent No. 6551795  
 ; GENERAL INFORMATION:  
 ; APPLICANT: MARC J. Rubenfield et al.  
 ; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO PSEUDOMONAS  
 ; TITLE OF INVENTION: AERUGINOSA FOR DIAGNOSTICS AND THERAPEUTICS  
 ; FILE REFERENCE: 107196.136  
 ; CURRENT APPLICATION NUMBER: US/09/252.991A  
 ; CURRENT FILING DATE: 1999-02-18  
 ; PRIOR APPLICATION NUMBER: US 60/074.788  
 ; PRIOR FILING DATE: 1998-02-18

; PRIOR APPLICATION NUMBER: US 60/094.190  
 ; PRIOR FILING DATE: 1998-07-27  
 ; NUMBER OF SEQ ID NOS: 33142  
 ; SEQ ID NO 27277  
 ; LENGTH: 401  
 ; TYPE: PRT  
 ; ORGANISM: Pseudomonas aeruginosa  
 US-09-252-991A-27277

Query Match 4.3%; Score 120; DB 4; Length 401;  
 Best Local Similarity 20.7%; Pred. No. 0.0034;  
 Matches 82; Conservative 53; Mismatches 153; Indels 108; Gaps 17;  
 177 PIKSGELKTKPTQHSVKELETTIGIQPDILICRSEQPIPASERRKIALFTVAAEKAVIS 236  
 81 PHIGNTG---TTPDAENRWMAAG-----LIIRDLPIASNNRKSQSLPDYKANGVVA 132  
 237 AIDADTIRIPLLRBGLD-----LVVDQLRVDPAADLS 273  
 133 IAGIDT--RRLTRILAEKSGONGCIIAGADATSERALBIARAEPGIGKNDLAKETVAERY 191  
 274 AMEKVVDGL---THPTDEVSIAVGRYVDHTDAYK-----SLNBAIHAQIHRHKVOIS 325  
 192 EKRSSVWNLSDSHP--RIP---AGELPYHVAVDYGVKALIRMLVARG-----CRIS 240  
 336 YIDSETIABGTAKLVNDAILVPGFGE--RGVEBKISTVPRARENKIPIYIGICLAKMS 383  
 241 VVPAQTTPASVYLA--LNPDIIFLSNGPDPEPCDYAIQAIREFLPTREIPIVFGICLHQLL 298  
 384 AVIEPARNVVGLGASTERFLPKSPHPIVGLITEMDBAGEIVTRDESDLGSTWELGAQ 443  
 299 ALASGAKTLKQGHGHGAN-----HPV-----QDLDS--GVVMTISQ 333  
 444 KCRILKADSLAFOLYQKDVITERHRRHYEFNNQYLKLEBAAGKFSGKSLDGLVLEIIEP 503  
 334 NHGFAVDSTL-----PDILRATHKSLPDGTIOGIERTKVAFSPFG----- 375  
 504 EHPWFLACQFHEPSTTPRNGHALFSGFVEAAAKH 539  
 376 -----HPEASPGHVDVAFIDRFTISAMARR 401

RESULT 10  
 US-08-116-098-2  
 ; Sequence 2, Application US/08116098  
 ; Patent No. 5428131  
 ; GENERAL INFORMATION:  
 ; APPLICANT: Trent, Jonathan D.  
 ; APPLICANT: Horwich, Arthur L.  
 ; TITLE OF INVENTION: Archaeobacterial Chaperonin-Mediated  
 ; TITLE OF INVENTION: Protein Stabilization  
 ; NUMBER OF SEQUENCES: 2  
 ; CORRESPONDENCE ADDRESS:  
 ; ADDRESSER: Kilpatrick & Cody  
 ; STREET: 1100 Peachtree Street  
 ; CITY: Atlanta  
 ; STATE: Georgia  
 ; COUNTRY: US  
 ; ZIP: 30309  
 ; COMPUTER READABLE FORM:  
 ; MEDIUM TYPE: Floppy disk  
 ; COMPUTER: IBM PC compatible  
 ; OPERATING SYSTEM: PC-DOS/MS-DOS  
 ; SOFTWARE: PatentIn Release #1.0, Version #1.25  
 ; CURRENT APPLICATION DATA:  
 ; APPLICATION NUMBER: US/08/116.098  
 ; FILING DATE:  
 ; CLASSIFICATION: 536  
 ; PRIOR APPLICATION DATA:  
 ; APPLICATION NUMBER: US 07/756627  
 ; FILING DATE: 09-SEP-1993  
 ; APPLICATION NUMBER: US 07/721974  
 ; FILING DATE: 27-JUN-1991

PRIOR APPLICATION DATA:  
 APPLICATION NUMBER: US 07/673158  
 FILING DATE: 18-MAR-1991  
 PRIOR APPLICATION DATA:  
 APPLICATION NUMBER: US 07/261573  
 FILING DATE: 24-OCT-1988  
 ATTORNEY/AGENT INFORMATION:  
 NAME: Padst, Patrea L.  
 REGISTRATION NUMBER: 31,284  
 REFERENCE/DOCKET NUMBER: YU102  
 TELECOMMUNICATION INFORMATION:  
 TELEPHONE: 404-815-6508  
 TELEFAX: 404-815-6555  
 INFORMATION FOR SEQ ID NO: 2:  
 SEQUENCE CHARACTERISTICS:  
 LENGTH: 552 amino acids  
 TYPE: amino acid  
 STRANDEDNESS: single  
 TOPOLOGY: linear  
 MOLECULE TYPE: protein  
 HYPOTHEICAL: NO  
 ANTI-SENSE: NO  
 FRAGMENT TYPE: N-terminal  
 ORIGINAL SOURCE:  
 ORGANISM: Sulfolobus shibatae  
 US-08-116-098-2

Query Match 4.0%; Score 111.5; DB 1; Length 552;  
 Best Local Similarity 21.1%; Pred. No. 0.04;  
 Matches 129; Conservative 73; Mismatches 176; Indels 233; Gaps 31;

5 IFITGVVSSLGKIAASSLAAL-LED-----RGL-KVTITLDPYINVDPGTMS 53  
 16 IILKSSSRITGKALANTAAVAKALEAKSTYGRGMDKMFVDSL----- 62  
 54 FQHGKVFYEDGAE-TDLDLGHYERFLKTYTKKNNFTTGQVYEGVLRNKKDYLGAT 111  
 63 ---GDITITNDGATILDKMDLQH-----PTGKLIVQIAKGQ----- 95  
 112 VQVIFHITDEIKRKYVESAGKQVALIEVGTVGDIESLPLETTIRQMGVELGRDALFI 171  
 96 -----DE-ETADGTTAVILAGELAKKAEDLLYKE----- 125  
 172 HLT-LVPYIKSAGELKTKPTQSHVSKELRTIGIOPDILICRSBOPIPASER---RKIALFT 227  
 126 HPTIIVSGYKKAELAKTIQ-----DI-----AQPVSINDTVLKKVAL-F 166  
 228 NVAEKAVISAIDMTYRIRPLIREGGLDLDVND-----QLRDVPAADLSAMB--KVV 279  
 167 SLGSKAVAGA-----REYLDLVKAVAGVAVELRGDKRYVDLDNVQIVKKG 212  
 280 DGLTHTDVSIAIVGKYVDHTDAYKSUNEALIHGIRHRYVQISYIDSEF----- 331  
 213 GGSINDTQVIGIVDKEVHHPMPRISNANI-ALLDASLEVEKEPELDAETRINDPTQM 271  
 332 ---IEAEGTAKLVNDAIIVPGG---FGRGVE-----GXISTVFARENKIPYLG 376  
 272 HKFLKEEENILMEKVDKIATGANNVITCGKIDENVAQHIAKGLIARAKGSLDEKDA 331  
 377 ICLGMD-----SAVIFAR---NVGLEGASTFELPKSGPHPIV-GLIT 416  
 332 RATGCGVISINIDELTSQDLGVAALVEERKVGEDKRVFVGACN---PXSVISILRGGLE 387  
 417 EWMDEAGELVTRDESDLGITVRLG-----AOKCRLKA-----DSLAPOLY 457  
 388 RYVDER-EKALADALCTADVTRDGRAVAGGAVEIENKRLKRAAPQVGGQGLAIYAY 446  
 458 QKDV-----ITERHRYEFNNQYLKOLEAAGMKFSGKSLDGRLVE 498  
 447 ANAIGLILMIAENAGLDPIDIKMQLRSIHENETKMYGLAN-----FTGN----- 492  
 499 IIELEHMPFL 509

DB 493 ----FEDMWKL 499

RESULT 11  
 US-08-687-590-32  
 Sequence 32, Application US/08687590  
 Patent No. 6255070  
 GENERAL INFORMATION:  
 APPLICANT: Wallison, Keith Robert  
 APPLICANT: Kubota, Hiroshi  
 APPLICANT: Ashworth, Alan  
 TITLE OF INVENTION: Folding Proteins  
 NUMBER OF SEQUENCES: 74  
 CORRESPONDENCE ADDRESS:  
 ADDRESSER: Townsend and Townsend and Crew LLP  
 STREET: Two Embarcadero Center, Eighth Floor  
 CITY: San Francisco  
 STATE: California  
 COUNTRY: United States of America  
 COMPUTER READABLE FORM:  
 MEDIUM TYPE: floppy disk  
 COMPUTER: IBM PC compatible  
 OPERATING SYSTEM: PC-DOS/MS-DOS  
 SOFTWARE: Patent Release #1.0, Version #1.25 (EPO)  
 CURRENT APPLICATION DATA:  
 APPLICATION NUMBER: US/08/687,590  
 FILING DATE: 31-JUL-1996  
 CLASSIFICATION: 435  
 PRIOR APPLICATION DATA:  
 APPLICATION NUMBER: PCT/GB95/00192  
 FILING DATE: 31-JAN-1995  
 PRIOR APPLICATION DATA:  
 APPLICATION NUMBER: GB 9401791.0  
 FILING DATE: 31-JAN-1994  
 PRIOR APPLICATION DATA:  
 APPLICATION NUMBER: GB 9418234.2  
 FILING DATE: 09-SEP-1994  
 ATTORNEY/AGENT INFORMATION:  
 NAME: Bastian, Kevin J.  
 REGISTRATION NUMBER: 34,774  
 REFERENCE/DOCKET NUMBER: 084619-000000US  
 INFORMATION FOR SEQ ID NO: 32:  
 SEQUENCE CHARACTERISTICS:  
 LENGTH: 552 amino acids  
 TYPE: amino acid  
 STRANDEDNESS: single  
 TOPOLOGY: linear  
 MOLECULE TYPE: peptide  
 US-08-687-590-32

Query Match 4.0%; Score 111.5; DB 3; Length 552;  
 Best Local Similarity 21.1%; Pred. No. 0.04;  
 Matches 129; Conservative 73; Mismatches 176; Indels 233; Gaps 31;

5 IFITGVVSSLGKIAASSLAAL-LED-----RGL-KVTITLDPYINVDPGTMS 53  
 16 IILKSSSRITGKALANTAAVAKALEAKSTYGRGMDKMFVDSL----- 62  
 54 FQHGKVFYEDGAE-TDLDLGHYERFLKTYTKKNNFTTGQVYEGVLRNKKDYLGAT 111  
 63 ---GDITITNDGATILDKMDLQH-----PTGKLIVQIAKGQ----- 95  
 112 VQVIFHITDEIKRKYVESAGKQVALIEVGTVGDIESLPLETTIRQMGVELGRDALFI 171  
 96 -----DE-ETADGTTAVILAGELAKKAEDLLYKE----- 125  
 172 HLT-LVPYIKSAGELKTKPTQSHVSKELRTIGIOPDILICRSBOPIPASER---RKIALFT 227  
 126 HPTIIVSGYKKAELAKTIQ-----DI-----AQPVSINDTVLKKVAL-F 166  
 228 NVAEKAVISAIDMTYRIRPLIREGGLDLDVND-----QLRDVPAADLSAMB--KVV 279  
 167 SLGSKAVAGA-----REYLDLVKAVAGVAVELRGDKRYVDLDNVQIVKKG 212

QY 280 DGLTHPTDEVSIAIVGKYVDHTDAVKSINLEAIHAGITRKRKVOISYDSEF----- 331  
 DB 213 GGSINDTOLVYGIIVDKVEVHPKMPRIENAKI-ALDASLEVEPDEDAERINDPTQM 271  
 QY 332 ---IABGPAKKNDAIIVPGG---FGRGVE-----GKSTVRFARENKIPYLG 376  
 DB 272 HKFLBEBENILKERVKDKLATGATNAVYTCCKGIDEVAGHTLAKKGLAVRRAKKSLEKLA 331  
 QY 377 ICLGNO-----SAVIEFAR-----NVVGEAGHTEFLPKSPHPVI-GHIT 416  
 DB 332 RAYGRVAINIDELISQDLGYAALVBERKVGEDKRVFEGAGN-----PKSVAILRGLE 387  
 QY 417 EMMDEAGELVTEDESDLGSTWELG-----AOKCKRKA-----DSLAPOLY 457  
 DB 388 RAVDET-ERALDPAIGTADVIRIDGSAVAGGAVEIEIAKRLKRYAPOVGKKEOLATIEAY 446  
 QY 458 OKDV-----ITERHRYEFNNQYLKQLEAAGMKFGSKSLDGRIVE 498  
 DB 447 ANAIEGLIMILANMGLDPIDKLMQRLSHENETKRYGLNL-----FTGN----- 492  
 QY 499 IIELEPHWEL 509  
 DB 493 ----PEDMWKL 499

## RESULT 12

US-09-066-046-2  
 ; Sequence 2, Application US/09066046A  
 ; Patent No. 6204252

## GENERAL INFORMATION:

APPLICANT: MURPHY, Cheryl  
 STOREY, James  
 COUGHLIN, Richard T.

TITLE OF INVENTION: CHARACTERIZATION OF GRANULOCYTIC  
 EHRILICHA AND METHODS OF USE

NUMBER OF SEQUENCES: 65  
 CORRESPONDENCE ADDRESS:  
 ADDRESSER: HALE AND DORR LLP  
 STREET: 60 State Street  
 CITY: Boston  
 STATE: MA  
 COUNTRY: United States  
 ZIP: 02109

## COMPUTER READABLE FORM:

MEDIUM TYPE: Floppy disk  
 COMPUTER: IBM PC compatible  
 OPERATING SYSTEM: PC-DOS/MS-DOS  
 SOFTWARE: Patent in Release #1.0, Version #1.30  
 CURRENT APPLICATION DATA:  
 APPLICATION NUMBER: US/09/066,046A  
 FILING DATE: 24-Apr-1998  
 CLASSIFICATION: <Unknown>

## ATTORNEY/AGENT INFORMATION:

NAME: Superko, Colleen  
 REGISTRATION NUMBER: 39,850  
 REFERENCE/DOCKET NUMBER: 106,941.155  
 TELECOMMUNICATION INFORMATION:  
 TELEPHONE: (617) 526-6000  
 TELEFAX: (617) 526-5000

## INFORMATION FOR SEQ ID NO: 2:

SEQUENCE CHARACTERISTICS:  
 LENGTH: 619 amino acids  
 TYPE: amino acid  
 STRANDEDNESS: single  
 TOPOLOGY: linear  
 MOLECULE TYPE: DNA (genomic)  
 HYPOTHETICAL: NO  
 ANTI-SENSE: NO

SEQUENCE DESCRIPTION: SEQ ID NO: 2:  
 US-09-066-046-2

Query Match 4.0%; Score 111.5; DB 3; Length 619;  
 Best Local Similarity 21.1%; Pred. No. 0.048;  
 Matches 95; Conservative 62; Mismatches 178; Indels 115; Gaps 20.

QY 44 INVDGTMSP-FOGEVFTEDGAEITDLIGHYERFLKTTTKKNTFTTGQVYQ----- 97  
 DB 223 VQVSSIIIVPESQTDVTEDETTITVDEGEYHFDIASGEHNDDLPALVLEADPTM 282  
 QY 98 VLNRERKGDYIGATVOYIPIHTDEIKRKYVE-----SABEK-----DVALIE 139  
 DB 283 LIANER-----SKTIESMP--SDLENDVOELGTLPLQGEVTSSEGTRESLPTDVSQDS 335  
 QY 140 VG-----GTVDIESLPLEET-IRQMGVEIGRDRALFIHTLVPIYKASGLK 186  
 DB 336 VGVSTDLAHSQEVEYTSSEVSTQDSLTNISQDSVGTSTL-LSHAKGVIESEG- 390  
 QY 187 TKPTQHSVK---ELRTGIQPDILICESEQPIPASEBKIALFTNVAEKAVISALDPTI 243  
 DB 391 ---TQDSLSADPINTVSESSTDLAHSQEVEYTSSEFTQDSLTNISQDSVGTSTLLEVH 447  
 QY 244 YRIPLLEBOGLDLYVDQLELDV--PAADLSAEKVYDGLTHPTD-----VSIAIVGK 296  
 DB 448 SQEVEYTSSEGTQDSLTNISQDSVGTSTDLAHSQEVEYTSSEFTQDSLTNISQDSVGT 507  
 QY 297 YVDHTDAVKSINLEAIHAGITRKRKVOISYDSEFTEIABGTAKKNVDAILVPGGFERG 356  
 DB 508 STD-LEVASGEVTSSEGTQDSLTNISQ--DSYGVSTDLAHSKGVITV-----SEGG 559  
 QY 357 VBKISTVRFARENKIPYLGICLQKQSVIVIFARNVYGLSEANSTFLPKSPH----- 409  
 DB 560 TQDSLSA-----DFPINTVSE--STDLEAHSPEGEIVSEV 593  
 QY 410 ----PVIGLITBMWDBAGELVTRDESDL 434  
 DB 594 STQDAPSTGVETIRFMD-----RDSDDV 616

## RESULT 13

US-09-328-352-7885  
 ; Sequence 7885, Application US/09328352  
 ; Patent No. 6562958

## GENERAL INFORMATION:

APPLICANT: Gary L. Breton et al.  
 TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO ACINETOBACTER  
 FILE REFERENCE: GTC99-03PA  
 CURRENT APPLICATION NUMBER: US/09/328,352  
 CURRENT FILING DATE: 1999-06-04  
 NUMBER OF SEQ ID NOS: 8252  
 SEQ ID NO 7885  
 LENGTH: 1407  
 TYPE: PRN  
 ORGANISM: Acinetobacter baumannii

US-09-328-352-7885

Query Match 3.9%; Score 109.5; DB 4; Length 1407;  
 Best Local Similarity 19.8%; Pred. No. 0.29;

Matches 139; Conservative 91; Mismatches 223; Indels 249; Gaps 38.

QY 41 DRYINVDGTMSPF-----ORGEVFTEDGAEI-----DIDLGH----- 74  
 DB 157 ESYVYTDG-MTFPEKYQLNDEEYFTALBEHGDFPAKKGABANVQDLKOIDLEABISR 215  
 QY 75 -YERFLKTTMTKK-----NN-----F 89  
 DB 216 LREEIPQTTSEKLLKAKSKRLKMEAFKDSNNKEMVMNVLPVLPDLRPLVPLEGRF 275  
 QY 90 TT---GVYEVYL-----RNE-----RKGYL-GAT 111  
 DB 276 ATSLNDLYIRKVINRNNRLKLLDEALPDIVYKNEKTLQESVDALLDNKGRGRATISGN 335  
 QY 112 VQVPHITDEIK--RKYESAEGKV-----ALIEVGTVDIE-SLPLETTRQMGVE 162

```

Db 336 KAPPLSLAMWIKQGRFRONLIGKRVDSGSSVITGPTLRLHQCGLP-----KMALE 390
QY 163 LGDRALFTHLTU-----VPYIKSAGELTKPTQHSVKELRTIGIOPDILICRSOPIPA 217
Db 391 LFPNP---FPAKLQASGQATTIKAAKQWERETPEVWDVLASVIRQHFWMLNRA-----PT 443
QY 218 SERKALALTNNAKKAVISALIDATYIRIPLIREGLDLDLVVDQLRLVP---AALISA 274
Db 444 LHLGLQAEPI-----LIEGKALRLHPLVCAAFNA-DPFGDQWAVHVPILTLERQLEA 495
QY 275 MEKVY-----DGLTPDEVSIAIVGKYVDHTDAYKSLNEALIIA----- 314
Db 496 RALMSTNNILSPANEPIIVSQDY---VLGIYITRAVNAKGEWFAADTHEVNAAL 552
QY 315 -----GHTRHXYQISYIDSETTEABGTAKLKNVDAILVPG-----GGEKGVGKISTV 364
Db 553 ATQGVAIHARVKR---VHQVTVINENGEBEQIIVDTTPGRCLLMEVVPBGILSDMTNL 609
QY 365 RFAREKRIPIYIGC---LGMOSAVIEFARNVVGLEGASHT-----FLPKSPHYV 411
Db 610 EMTKKNISKLINSCYKRLGIKDTVI-FADQLATLIGFRQATRSVSIVGMEDMLIPETKTI 668
QY 412 I-----GITTEWMDAGELVTRDESDLGTMRLGAQKCRLLKADSLAQ 455
Db 669 IDKATEVEIEEQFROGFVT---AGERYNKVVDI---WARTVDQAKAMMDNLSTY 719
QY 456 LYO-KVYTERHRRHRENNQY-----LKOLEAAGMKRSGSLDGRLEVI-I 500
Db 720 LVNKKQGEDEKQK---SPNSIYMSDSGARSAQIRU-AGMRLLMAKPDGSIITETPI 774
QY 501 ELBEHPWFLACQFHEPFTSTPRNGHALFSGFVZAANKHTQG 542
Db 775 KANPREGIYLOQ---FIST---HGARKGLADTALKTANSG 809

```

```

RESULT 14
US-09-252-991A-32742
; Sequence 32742, Application US/09252991A
; Patent No. 6551795
; GENERAL INFORMATION:
; APPLICANT: Marc J. Rubenfield et al.
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO PSEUDOMONAS
; FILE REFERENCE: 107196.136
; CURRENT APPLICATION NUMBER: US/09/252,991A
; PRIOR FILING DATE: 1999-02-18
; PRIOR APPLICATION NUMBER: US 60/074,788
; PRIOR FILING DATE: 1998-02-18
; PRIOR APPLICATION NUMBER: US 60/094,190
; NUMBER OF SEQ ID NOS: 33142
; SEQ ID NO 32742
; LENGTH: 809
; TYPE: PRT
; ORGANISM: Pseudomonas aeruginosa
US-09-252-991A-32742

```

```

Query Match 3.9%; Score 109; DB 4; Length 809;
Best Local Similarity 19.2%; Pred. No. 0.13;
Matches 126; Conservative 74; Mismatches 203; Indels 252; Gaps 31;
QY 42 PYINDRCMTSPFQHGIVFTEDEA-----ETDLDLGHVERPLKTTMTKNNFTTGQYV 95
Db 19 PGVAVAGAPPIVAHVVQPTLDAVVAITQGPVDLHLRQRKRPDRDHPQVNRQRF 78
QY 96 -----EQLVNRKSG-DYLGATVQVAPHTIDEIKRRV-----YESAKQV 135
Db 79 GQRTDVLAAQALGEGAGAPFVGVDVDAALAHQLAETLSAVGNPFDVAAVADAVREVRV 138
QY 136 ALIE-----VGGTVGD-----ISLPLETTRQWG---VELGRBALFIHTLVP 177
Db 139 AFAEQPAAIAGEVABGLDAHRLAVVEDRHVLAQHRLGSKQVAILARQAV----- 190

```

```

QY 178 YIKSAGELTKPT-----QHSVKELRTIGIOPDIL----- 207
Db 191 -----GQYQVDPSSGGVQAPPAVQAHNHLDPVAQRAEQPLHQRRQEPDILAAVAPGDR 245
QY 208 -----ICRSEQPIPASERRKIALF-TNVAEKA---VISAIDATIYRIPLL 250
Db 246 VYRIYALADGIRSLCQAQ-Q-PSABRQPAHLSGVAGRGEQVAGAVDQ----- 295
QY 251 REQGLDLDLVQDLRLVPADLS---AMEKYVDGLTHTDEVSIAIVGKYVDHTDAYKSLN 308
Db 296 -----DQLFGRVVAQLATQADQHYDQ-----ALAQPLVAD----- 329
QY 309 EALIHAGIHTHKKQISYIDSETTEABGTAKLKNVDAILVPGGGEKGVGKISTVFEAR 368
Db 330 -LVH-----QVAAQMRVAVDSABQLEFPAAGEQCFVARKV----- 365
QY 369 ENKIPYIGICLGMOSAVIEFARNVVGLEGA-----HSTELPKSPHYV 413
Db 366 DQAP---VGVQPVAVELPEGLPBGAGLGMANPADQILRPHQFARVQRDLHVIVG 420
QY 414 LITTEMDAGELVTRDESD---LGTMRLGAQKCRLLKADSLAQVQKDYITE----- 464
Db 421 AALQ-ADDAIDLVAAPBDQDPLHARHQLAGQCAVLAGQADVQHQKDLALAMGALGX 479
QY 465 -----HRR---HRYEFNNQYL-----KOLEAAGMKF 487
Db 480 AGFRLGVAGABDVIALIAIEIGLEKLTNRVYVDHQVARNRHHILAPAGRGVSCSGLSD 539
QY 488 SKSLDGLVAILIELPEHPWFL---ACQFH-PEFTSTPR-----NGHAL 527
Db 540 GGRKPGRL-----PLPAAVPLNGRSTSFHRTPRAMAGARLQAIHPLANTHAL 550

```

```

RESULT 15
US-09-252-991A-21763
; Sequence 21763, Application US/09252991A
; Patent No. 6551795
; GENERAL INFORMATION:
; APPLICANT: Marc J. Rubenfield et al.
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO PSEUDOMONAS
; FILE REFERENCE: 107196.136
; CURRENT APPLICATION NUMBER: US/09/252,991A
; PRIOR FILING DATE: 1999-02-18
; PRIOR APPLICATION NUMBER: US 60/074,788
; PRIOR FILING DATE: 1998-02-18
; PRIOR APPLICATION NUMBER: US 60/094,190
; NUMBER OF SEQ ID NOS: 33142
; SEQ ID NO 21763
; LENGTH: 1700
; TYPE: PRT
; ORGANISM: Pseudomonas aeruginosa
US-09-252-991A-21763

```

```

Query Match 3.9%; Score 107.5; DB 4; Length 1700;
Best Local Similarity 22.0%; Pred. No. 0.64; Indels 97; Gaps 16;
Matches 84; Conservative 43; Mismatches 156;
QY 110 ATVCYPIPHITDEIKRRVYESABKQVALIEVGTVG---DIESLPLETTRQMGVELGR 165
Db 534 AVEVEVGR-----REVVPRAQGVAVQVMAAHQHLVAALDQRTVGCOPTLRLGQVDHRR 588
QY 166 DRALFIHTVPIYISABE-----LKTPTQHSVELRTIGIOP-DI 206
Db 569 EDLAVHLDLPQHDVAVQQRHLRAQHTQLQVQLVVSQGVNHQVAFQRLVAAQPLDV 648
QY 207 LICR-----SEQPIPASERRKIALFTNVAEKAVIS---ALDATTYRIPLL---RE 252
Db 649 ALAGARHHRVADQPL-----LVETVAEALTLASVNHQAQAEQVAVREBELQFGERR 699
QY 253 QGLDVLVDQLRLVPAADLSAMKRVNDGLTHTPDEVSIAIVGKYVDHTDAYKSLNEALI 312

```



```

Db      700 VGLDQVIVAGARR-----GVAGHALDPGDARQATERRAA 733
Qy      313 HAGIHTRHKKVOISYIDSETEAEQT-----AKLKNVDAILVPGGPG--ERGVE-GKIST 363
Db      734 ALTLIDVGRASQA--VDSHALALPCTGIVAGALLADRDATATGAVGNGTERGAATVIGSE 791
Qy      364 VRFARENKIPYLGICLMQOSAVIEFARNVVG--LEGHSTETPLPKSPHPVIGLITEMMDE 421
Db      792 ARAETEOAVVTLGQANGQGNRYDLAPRGVGRQLRGQRHDERNPAAAHRAVAHILAAADE 851
Qy      422 ---AGELVTRDE---DSDIGG 436
Db      852 GVAGAGDAVVDARAGIDRDVAG 873

```

Search completed: January 29, 2004, 15:57:07  
 Job time : 18.0734 secs

GenCore version 5.1.6  
Copyright (c) 1993 - 2004 Compugen Ltd.

OM protein - protein search, using sw model

Run on: January 29, 2004, 15:54:19 : Search time 37.1346 Seconds  
(without alignments)  
3044.503 Million cell updates/sec

Title: US-09-941-947a-16

Perfect score: 2788  
Sequence: 1 MKXFFITGVVSSIGKXIA.....HALFGFVEAAKKTGGTA 544

Scoring table: BLOSUM62  
Gapop 10.0, Gapext 0.5

Searched: 789580 seqs, 207824079 residues

Total number of hits satisfying chosen parameters: 789580

Minimum DB seq length: 0  
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : Published Applications AA:\*

- 1: /cgn2\_6/ptodata/1/pubpaa/US07\_PUBCOMB.pep.\*
- 2: /cgn2\_6/ptodata/1/pubpaa/PCT\_NEW\_PUB.pep.\*
- 3: /cgn2\_6/ptodata/1/pubpaa/US06\_NEW\_PUB.pep.\*
- 4: /cgn2\_6/ptodata/1/pubpaa/US06\_PUBCOMB.pep.\*
- 5: /cgn2\_6/ptodata/1/pubpaa/US07\_NEW\_PUB.pep.\*
- 6: /cgn2\_6/ptodata/1/pubpaa/PCT05\_PUBCOMB.pep.\*
- 7: /cgn2\_6/ptodata/1/pubpaa/US08\_NEW\_PUB.pep.\*
- 8: /cgn2\_6/ptodata/1/pubpaa/US08\_PUBCOMB.pep.\*
- 9: /cgn2\_6/ptodata/1/pubpaa/US09\_PUBCOMB.pep.\*
- 10: /cgn2\_6/ptodata/1/pubpaa/US09B\_PUBCOMB.pep.\*
- 11: /cgn2\_6/ptodata/1/pubpaa/US09C\_PUBCOMB.pep.\*
- 12: /cgn2\_6/ptodata/1/pubpaa/US09\_NEW\_PUB.pep.\*
- 13: /cgn2\_6/ptodata/1/pubpaa/US10A\_PUBCOMB.pep.\*
- 14: /cgn2\_6/ptodata/1/pubpaa/US10B\_PUBCOMB.pep.\*
- 15: /cgn2\_6/ptodata/1/pubpaa/US10C\_PUBCOMB.pep.\*
- 16: /cgn2\_6/ptodata/1/pubpaa/US10\_NEW\_PUB.pep.\*
- 17: /cgn2\_6/ptodata/1/pubpaa/US60\_NEW\_PUB.pep.\*
- 18: /cgn2\_6/ptodata/1/pubpaa/US60\_PUBCOMB.pep.\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

## SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	2788	100.0	544	10	US-09-934-903-12
2	2788	100.0	544	10	US-09-934-903-12
3	2788	100.0	544	11	US-09-941-947a-16
4	2008	72.0	542	9	US-09-815-242-11933
5	1916	68.7	545	9	US-09-815-242-10275
6	1908.5	68.5	545	9	US-09-815-242-11175
7	1783	64.0	511	9	US-09-815-242-11365
8	1576	56.5	536	9	US-09-815-242-12372
9	1576	56.5	536	9	US-09-815-242-12378
10	1569	56.3	536	9	US-09-815-242-4984
11	1569	56.3	536	9	US-09-815-242-10727
12	1561	56.0	535	9	US-09-815-242-13376
13	1444.5	51.8	538	9	US-09-815-242-11320
14	1432.5	51.4	549	15	US-10-156-761-14031
15	1372	49.2	554	10	US-09-738-626-5066

16	1355.5	48.6	537	12	US-10-289-762-246	Sequence 246, App
17	1006.5	36.1	1003	12	US-10-161-051-91	Sequence 91, App
18	513	18.4	199	12	US-10-264-049-2881	Sequence 2881, App
19	353.5	12.7	199	15	US-10-106-698-4980	Sequence 4980, App
20	345.5	12.4	195	9	US-09-925-237-773	Sequence 773, App
21	327.5	11.7	94	12	US-09-864-408A-3914	Sequence 3914, App
22	245	8.8	135	12	US-09-864-408A-2350	Sequence 2350, App
23	204	7.3	135	12	US-10-264-049-4217	Sequence 4217, App
24	126.5	4.5	238	12	US-10-314-657-34	Sequence 34, App
25	126	4.5	374	12	US-10-369-493-10131	Sequence 10131, App
26	125.5	4.5	1390	12	US-10-369-493-11073	Sequence 11073, App
27	124	4.4	374	12	US-10-369-493-15844	Sequence 15844, App
28	124	4.4	374	12	US-10-369-493-16223	Sequence 16223, App
29	123.5	4.4	371	12	US-10-369-493-29	Sequence 29, App
30	122	4.4	845	10	US-09-738-626-4347	Sequence 4347, App
31	120	4.3	382	12	US-10-369-493-691	Sequence 691, App
32	120	4.3	382	12	US-10-369-493-15474	Sequence 15474, App
33	120	4.3	382	16	US-10-210-115-33	Sequence 33, App
34	116.5	4.2	656	12	US-10-369-493-18670	Sequence 18670, App
35	115.5	4.1	650	12	US-10-369-493-1338	Sequence 1338, App
36	115.5	4.1	650	12	US-10-369-493-20401	Sequence 20401, App
37	115.5	4.1	1687	14	US-10-094-679-3	Sequence 3, App
38	114	4.1	1071	12	US-10-369-493-23117	Sequence 23117, App
39	113.5	4.1	106	12	US-09-864-408A-2256	Sequence 2256, App
40	113	4.1	362	12	US-10-369-493-17332	Sequence 17332, App
41	112.5	4.0	371	12	US-10-369-493-11567	Sequence 11567, App
42	112.5	4.0	537	12	US-10-369-493-89	Sequence 89, App
43	111.5	4.0	728	12	US-10-369-493-15429	Sequence 15429, App
44	111.5	4.0	728	12	US-10-369-493-15797	Sequence 15797, App
45	111.5	4.0	728	12	US-10-369-493-16179	Sequence 16179, App

## ALIGNMENTS

RESULT 1  
US-09-934-903-12  
Sequence 12, Application US/09934903  
Patent No. US20020102690A1  
GENERAL INFORMATION:  
APPLICANT: Kofas, Mattheos  
APPLICANT: Odem, J. Martin  
APPLICANT: Schenzle, Andreas J.  
APPLICANT: No. US20020102690A1ton, Kelley C.  
APPLICANT: Tomb, Jean-Francois  
APPLICANT: Roviére, Pierre  
APPLICANT: Picataggio, Stephen  
APPLICANT: Cheng, Qiong  
TITLE OF INVENTION: Genes Involved in Isoprenoid Compounds Production  
FILE REFERENCE: C1646 US NA  
CURRENT APPLICATION NUMBER: US/09/934, 903  
CURRENT FILING DATE: 2001-08-22  
PRIOR APPLICATION NUMBER: 60/229,907  
PRIOR FILING DATE: September 1, 2001  
NUMBER OF SEQ ID NOS: 24  
SOFTWARE: Microsoft Office 97  
SEQ ID NO 12  
LENGTH: 544  
TYPE: PRT  
ORGANISM: Methylobionas 16a  
FEATURE:  
OTHER INFORMATION: Amino acid sequences encoded by ORF6  
US-09-934-903-12

Query Match 100.0%; Score 2788; DB 10; Length 544;  
Best Local Similarity 100.0%; Pred. No. 1.5e-254;  
Matches 544; Conservative 0; Mismatches 0; Gaps 0;

QY 1 MKXFFITGVVSSIGKXIAAIAIEDGKLVTKADPYINVDGTSPPQHGVSF 60  
|||||  
DB 1 MKXFFITGVVSSIGKXIAAIAIEDGKLVTKADPYINVDGTSPPQHGVSF 60  
|||||  
QY 61 VTDEGATLDLDIGHYERLTKTKKNNFTTGQVYEQVLRNRRGDIYATVQVPHITD 120  
|||||

Db 61 VTGSAETDLDLGHIERLKTMTKNNFTTGQVYEGVLARNERGDYIGATVQVPIPHITD 120  
 QY 121 EIKRRVYESAGKVALIEVGTVGDIESTLPLETTIRMGVGLGRDALFHLTLVPYIK 180  
 Db 121 EIKRRVYESAGKVALIEVGTVGDIESTLPLETTIRMGVGLGRDALFHLTLVPYIK 180  
 QY 181 SAGELKTPQSHVSKELRTIGIOPDILICRSEOPIPASERKIALFTNVAEKAVISIDA 240  
 Db 181 SAGELKTPQSHVSKELRTIGIOPDILICRSEOPIPASERKIALFTNVAEKAVISIDA 240  
 QY 241 DTIYRIPLLREOGIDLDLVQDLRLDVPADLSAMKRVNDGLTHPTDEVSIAIVGYVDH 300  
 Db 241 DTIYRIPLLREOGIDLDLVQDLRLDVPADLSAMKRVNDGLTHPTDEVSIAIVGYVDH 300  
 QY 301 TDAYSINLEALIHAGIHTHRKVQISYIDSETTIEAGTAKIKNDAILVPGFGEERGVGK 360  
 Db 301 TDAYSINLEALIHAGIHTHRKVQISYIDSETTIEAGTAKIKNDAILVPGFGEERGVGK 360  
 QY 361 ISTVRFARENKIPLYIGICMGOSAVIEFARNVGLGASTETPLPKSPHPVIGLITEMD 420  
 Db 361 ISTVRFARENKIPLYIGICMGOSAVIEFARNVGLGASTETPLPKSPHPVIGLITEMD 420  
 QY 421 EAGELVTRDESDLDGGTMRGAKCKRLKADSLAFOLYQKDVITERHRHRYEFNNQYIKOL 480  
 Db 421 EAGELVTRDESDLDGGTMRGAKCKRLKADSLAFOLYQKDVITERHRHRYEFNNQYIKOL 480  
 QY 481 EAGMKFSGKSLDGLRVEIIELEPHWFLACQHPHFTSTPRNGHALFSGFVEAAAKHKT 540  
 Db 481 EAGMKFSGKSLDGLRVEIIELEPHWFLACQHPHFTSTPRNGHALFSGFVEAAAKHKT 540  
 QY 541 QGTA 544  
 Db 541 QGTA 544

## RESULT 2

US-09-934-868-70  
 ; Sequence 70, Application US/09934868  
 ; Patent No. US2002017190A1  
 ; GENERAL INFORMATION:  
 ; APPLICANT: Odom, James M.  
 ; APPLICANT: Koffas, Mattheos  
 ; APPLICANT: Schenzzle, Andreas J.  
 ; TITLE OF INVENTION: DENTRIFYING METHANOTROPHIC BACTERIAL STRAIN  
 ; FILE REFERENCE: CL1596 US NA  
 ; CURRENT APPLICATION NUMBER: US/09/934,868  
 ; CURRENT FILING DATE: 2001-08-22  
 ; PRIOR APPLICATION NUMBER: 60/229,858  
 ; PRIOR FILING DATE: 2000-09-01  
 ; NUMBER OF SEQ ID NOS: 81  
 ; SOFTWARE: Microsoft Office 97  
 ; SEQ ID NO 70  
 ; LENGTH: 544  
 ; TYPE: PRF  
 ; ORGANISM: Methylomonas 16a  
 ; OTHER INFORMATION: Amino acid sequences encoded by ORF6 - PRNG  
 ; US-09-934-868-70

Query Match 100.0%; Score 2788; DB 10; Length 544;  
 Best Local Similarity 100.0%; Pred. No. 1.5e-254;  
 Matches 544; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 MTKXIFITGVVSSLGKGIASSLAAILBDRGLKVTITKLDPIYINVDGTMSPFOHGEVF 60  
 Db 1 MTKXIFITGVVSSLGKGIASSLAAILBDRGLKVTITKLDPIYINVDGTMSPFOHGEVF 60  
 QY 61 VTGSAETDLDLGHIERLKTMTKNNFTTGQVYEGVLARNERGDYIGATVQVPIPHITD 120  
 Db 61 VTGSAETDLDLGHIERLKTMTKNNFTTGQVYEGVLARNERGDYIGATVQVPIPHITD 120  
 QY 121 EIKRRVYESAGKVALIEVGTVGDIESTLPLETTIRMGVGLGRDALFHLTLVPYIK 180

Db 121 EIKRRVYESAGKVALIEVGTVGDIESTLPLETTIRMGVGLGRDALFHLTLVPYIK 180  
 QY 181 SAGELKTPQSHVSKELRTIGIOPDILICRSEOPIPASERKIALFTNVAEKAVISIDA 240  
 Db 181 SAGELKTPQSHVSKELRTIGIOPDILICRSEOPIPASERKIALFTNVAEKAVISIDA 240  
 QY 241 DTIYRIPLLREOGIDLDLVQDLRLDVPADLSAMKRVNDGLTHPTDEVSIAIVGYVDH 300  
 Db 241 DTIYRIPLLREOGIDLDLVQDLRLDVPADLSAMKRVNDGLTHPTDEVSIAIVGYVDH 300  
 QY 301 TDAYSINLEALIHAGIHTHRKVQISYIDSETTIEAGTAKIKNDAILVPGFGEERGVGK 360  
 Db 301 TDAYSINLEALIHAGIHTHRKVQISYIDSETTIEAGTAKIKNDAILVPGFGEERGVGK 360  
 QY 361 ISTVRFARENKIPLYIGICMGOSAVIEFARNVGLGASTETPLPKSPHPVIGLITEMD 420  
 Db 361 ISTVRFARENKIPLYIGICMGOSAVIEFARNVGLGASTETPLPKSPHPVIGLITEMD 420  
 QY 421 EAGELVTRDESDLDGGTMRGAKCKRLKADSLAFOLYQKDVITERHRHRYEFNNQYIKOL 480  
 Db 421 EAGELVTRDESDLDGGTMRGAKCKRLKADSLAFOLYQKDVITERHRHRYEFNNQYIKOL 480  
 QY 481 EAGMKFSGKSLDGLRVEIIELEPHWFLACQHPHFTSTPRNGHALFSGFVEAAAKHKT 540  
 Db 481 EAGMKFSGKSLDGLRVEIIELEPHWFLACQHPHFTSTPRNGHALFSGFVEAAAKHKT 540  
 QY 541 QGTA 544  
 Db 541 QGTA 544

## RESULT 3

US-09-941-947A-16  
 ; Sequence 16, Application US/09941947A  
 ; Publication No. US20030003528A1  
 ; GENERAL INFORMATION:  
 ; APPLICANT: Brzostowicz, Patricia C.  
 ; APPLICANT: Cheng, Qiong  
 ; APPLICANT: DiCosimo, Deana J.  
 ; APPLICANT: Koffas, Mattheos  
 ; APPLICANT: Miller, Edward S. Jr.  
 ; APPLICANT: Odom, J. Martin  
 ; APPLICANT: Picataggio, Steve  
 ; APPLICANT: Rouviere, Pierre E.  
 ; TITLE OF INVENTION: CAROTENOID PRODUCTION FROM A SINGLE CARBON SOURCE  
 ; FILE REFERENCE: CL1903 US NA  
 ; CURRENT APPLICATION NUMBER: US/09/941,947A  
 ; CURRENT FILING DATE: 2001-09-01  
 ; PRIOR APPLICATION NUMBER: 60/229,907  
 ; PRIOR FILING DATE: 2000-09-01  
 ; PRIOR APPLICATION NUMBER: 60/229,858  
 ; PRIOR FILING DATE: 2000-09-01  
 ; NUMBER OF SEQ ID NOS: 60  
 ; SOFTWARE: Microsoft Office 97  
 ; SEQ ID NO 16  
 ; LENGTH: 544  
 ; TYPE: PRF  
 ; ORGANISM: Methylomonas 16a  
 ; US-09-941-947A-16

Query Match 100.0%; Score 2788; DB 11; Length 544;  
 Best Local Similarity 100.0%; Pred. No. 1.5e-254;  
 Matches 544; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 MTKXIFITGVVSSLGKGIASSLAAILBDRGLKVTITKLDPIYINVDGTMSPFOHGEVF 60  
 Db 1 MTKXIFITGVVSSLGKGIASSLAAILBDRGLKVTITKLDPIYINVDGTMSPFOHGEVF 60  
 QY 61 VTGSAETDLDLGHIERLKTMTKNNFTTGQVYEGVLARNERGDYIGATVQVPIPHITD 120  
 Db 61 VTGSAETDLDLGHIERLKTMTKNNFTTGQVYEGVLARNERGDYIGATVQVPIPHITD 120

```

QY 121 EIKRKYVESAEKQVALLIVGGTVDIESLPLEFTIRQMGVLEGRDALFTHTLVPIYK 180
DB 121 EIKRKYVESAEKQVALLIVGGTVDIESLPLEFTIRQMGVLEGRDALFTHTLVPIYK 180
QY 181 SAGELEKTKPTQHSYKELRTIGIQPDILICRSEOPIPASERKIALFTVAEKAIVSAIDA 240
DB 181 SAGELEKTKPTQHSYKELRTIGIQPDILICRSEOPIPASERKIALFTVAEKAIVSAIDA 240
QY 241 DTIYRIPILLREOGDLVDLVDQRLDVPADLSAMEKVVDELTPDEVSIAIVKQYVDH 300
DB 241 DTIYRIPILLREOGDLVDLVDQRLDVPADLSAMEKVVDELTPDEVSIAIVKQYVDH 300
QY 301 TDVAYSINBALIHAGIHTHRKVQISYIDSETIIEAGSTALKNVDAIIVPGGFGGEGVGEK 360
DB 301 TDVAYSINBALIHAGIHTHRKVQISYIDSETIIEAGSTALKNVDAIIVPGGFGGEGVGEK 360
QY 361 ISTVFARENKIPYLGICLQMSAVIEFARNVVLGEGASTEFLEPKSPHPVIGLITTEMD 420
DB 361 ISTVFARENKIPYLGICLQMSAVIEFARNVVLGEGASTEFLEPKSPHPVIGLITTEMD 420
QY 421 ENGELVTRDESDLGSTMRLEAGCKRLKADSLAFQLYQKDVITERRHRRHYEFNNQYKQL 480
DB 421 ENGELVTRDESDLGSTMRLEAGCKRLKADSLAFQLYQKDVITERRHRRHYEFNNQYKQL 480
QY 481 EAAGKKSCKSLDGRLEVEIIELEPEHPWFLACQFHEPFTSPRNGHALFSGFVEAAAKKH 540
DB 481 EAAGKKSCKSLDGRLEVEIIELEPEHPWFLACQFHEPFTSPRNGHALFSGFVEAAAKKH 540
QY 541 OGTA 544
DB 541 OGTA 544

RESULT 4
US-09-815-242-11933
; Sequence 11933, Application US/09815242
; Patent No. US20020061569A1
; GENERAL INFORMATION:
; APPLICANT: Haselbeck, Robert
; APPLICANT: Ohlsen, Karl L.
; APPLICANT: Zyskind, Judith W.
; APPLICANT: Wall, Daniel
; APPLICANT: Treawick, John D.
; APPLICANT: Carr, Grant J.
; APPLICANT: Yamamoto, Robert T.
; APPLICANT: Xu, H. Howard
; TITLE OF INVENTION: Identification of Essential Genes in
; FILE REFERENCE: ELITRA.011A
; CURRENT APPLICATION NUMBER: US/09/815,242
; PRIOR FILING DATE: 2001-03-21
; PRIOR APPLICATION NUMBER: 60/191,078
; PRIOR FILING DATE: 2000-03-21
; PRIOR APPLICATION NUMBER: 60/206,848
; PRIOR FILING DATE: 2000-05-23
; PRIOR APPLICATION NUMBER: 60/207,727
; PRIOR FILING DATE: 2000-05-26
; PRIOR APPLICATION NUMBER: 60/242,578
; PRIOR FILING DATE: 2000-10-23
; PRIOR APPLICATION NUMBER: 60/253,625
; PRIOR FILING DATE: 2000-11-27
; PRIOR APPLICATION NUMBER: 60/257,931
; PRIOR FILING DATE: 2000-12-22
; PRIOR APPLICATION NUMBER: 60/269,308
; PRIOR FILING DATE: 2001-02-16
; NUMBER OF SEQ ID NOS: 14110
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO: 11933
; LENGTH: 542
; TYPE: PKT
; ORGANISM: Pseudomonas aeruginosa
US-09-815-242-11933

```

```

Query Match 72.0%, Score 2008, DB 9, Length 542;
Best Local Similarity 71.2%, Pred. No. 8,56-181;
Matches 383, Conservative 64, Mismatches 91, Indels 0, Gaps 0;

QY 1 MTKGIFITGGVSSLGKGIASSLAATIEDRLKVTCTKDPYINVDGTMSPFOHGEVF 60
DB 1 MTKGIFITGGVSSLGKGIASSLAATIEDRLKVTCTKDPYINVDGTMSPFOHGEVF 60
QY 61 VTEDGAEITDLDGHERPLKTKTKNNFTTGQYVEQYLRNERKQDYIACVAVIPIHTD 120
DB 61 VTEDGAEITDLDGHERPLKTKTKNNFTTGQYVEQYLRNERKQDYIACVAVIPIHTD 120
QY 121 EIKRKYVESAEKQVALLIVGGTVDIESLPLEFTIRQMGVLEGRDALFTHTLVPIYK 180
DB 121 EIKRKYVESAEKQVALLIVGGTVDIESLPLEFTIRQMGVLEGRDALFTHTLVPIYK 180
QY 181 SAGELEKTKPTQHSYKELRTIGIQPDILICRSEOPIPASERKIALFTVAEKAIVSAIDA 240
DB 181 SAGELEKTKPTQHSYKELRTIGIQPDILICRSEOPIPASERKIALFTVAEKAIVSAIDA 240
QY 241 DTIYRIPILLREOGDLVDLVDQRLDVPADLSAMEKVVDELTPDEVSIAIVKQYVDH 300
DB 241 DTIYRIPILLREOGDLVDLVDQRLDVPADLSAMEKVVDELTPDEVSIAIVKQYVDH 300
QY 301 TDVAYSINBALIHAGIHTHRKVQISYIDSETIIEAGSTALKNVDAIIVPGGFGGEGVGEK 360
DB 301 TDVAYSINBALIHAGIHTHRKVQISYIDSETIIEAGSTALKNVDAIIVPGGFGGEGVGEK 360
QY 361 ISTVFARENKIPYLGICLQMSAVIEFARNVVLGEGASTEFLEPKSPHPVIGLITTEMD 420
DB 361 ISTVFARENKIPYLGICLQMSAVIEFARNVVLGEGASTEFLEPKSPHPVIGLITTEMD 420
QY 421 ENGELVTRDESDLGSTMRLEAGCKRLKADSLAFQLYQKDVITERRHRRHYEFNNQYKQL 480
DB 421 ENGELVTRDESDLGSTMRLEAGCKRLKADSLAFQLYQKDVITERRHRRHYEFNNQYKQL 480
QY 481 EAAGKKSCKSLDGRLEVEIIELEPEHPWFLACQFHEPFTSPRNGHALFSGFVEAAAKKH 538
DB 481 EAAGKKSCKSLDGRLEVEIIELEPEHPWFLACQFHEPFTSPRNGHALFSGFVEAAAKKH 538

RESULT 5
US-09-815-242-10275
; Sequence 10275, Application US/09815242
; Patent No. US20020061569A1
; GENERAL INFORMATION:
; APPLICANT: Haselbeck, Robert
; APPLICANT: Ohlsen, Karl L.
; APPLICANT: Zyskind, Judith W.
; APPLICANT: Wall, Daniel
; APPLICANT: Treawick, John D.
; APPLICANT: Carr, Grant J.
; APPLICANT: Yamamoto, Robert T.
; APPLICANT: Xu, H. Howard
; TITLE OF INVENTION: Identification of Essential Genes in
; FILE REFERENCE: ELITRA.011A
; CURRENT APPLICATION NUMBER: US/09/815,242
; PRIOR FILING DATE: 2001-03-21
; PRIOR APPLICATION NUMBER: 60/191,078
; PRIOR FILING DATE: 2000-03-21
; PRIOR APPLICATION NUMBER: 60/206,848
; PRIOR FILING DATE: 2000-05-23
; PRIOR APPLICATION NUMBER: 60/207,727
; PRIOR FILING DATE: 2000-05-26
; PRIOR APPLICATION NUMBER: 60/242,578
; PRIOR FILING DATE: 2000-10-23
; PRIOR APPLICATION NUMBER: 60/253,625
; PRIOR FILING DATE: 2000-11-27
; PRIOR APPLICATION NUMBER: 60/257,931
; PRIOR FILING DATE: 2000-12-22
; PRIOR APPLICATION NUMBER: 60/269,308
; PRIOR FILING DATE: 2001-02-16

```

NUMBER OF SEQ ID NOS: 14110  
 SOFTWARE: FastSeq for windows version 4.0  
 SEQ ID NO 10275  
 LENGTH: 545  
 TYPE: PRT  
 ORGANISM: Escherichia coli  
 US-09-815-242-10275

Query Match 68.7%; Score 1916; DB 9; Length 545;  
 Best Local Similarity 67.7%; Pred. No. 4.3e-172;  
 Matches 363; Conservative 71; Mismatches 102; Indels 0; Gaps 0;

QY 2 TKRFTITGVVSSLSGKGIASSLAALIEDRLKVTITLDPYINVDPGTMSPFQHGVEV 61  
 DB 3 TNYIFVTGVVSSLSGKGIASSLAALIEDRLKVTITLDPYINVDPGTMSPFQHGVEV 62  
 QY 62 TEDGATDLDLGHYERFLKTTMTKKNFTTQVYBOVLNERNKGDYLGATVQVPIHITDE 121  
 DB 63 TEDGATDLDLGHYERFLKTTMTKKNFTTQVYBOVLNERNKGDYLGATVQVPIHITDE 122  
 QY 122 IKRRVYESAEKDVALLIEVGTVGDIESLPLEFTIRQMGVELGRDRLAFHILTLVPIYKS 181  
 DB 123 IKRRVYESAEKDVALLIEVGTVGDIESLPLEFTIRQMGVELGRDRLAFHILTLVPIYKS 182  
 QY 182 AGELTKPTQHSVKELRTTIGIOPDILICRSEOPTPASBRKIALFTVAEKAVISAIDAD 241  
 DB 183 SEEVTKPTQHSVKELRTTIGIOPDILICRSDRAVPNERAKIALFCNVERAKVISLKYVD 242  
 QY 242 TIYRIPLLEKQGLDDLVVDQLRLDVPADLSAMEKVVYDGLTHPTDEVSIAIVKTYDHT 301  
 DB 243 SIYKIPGLKSGQLDDVYICRFSLNCPEANLSEWQVIFEEANPVSEVITIGWVKYIELP 302  
 QY 302 DAKSLNELLHAGHTRHVKQISYIDSETEABGTALKNVDAILVYGGFGEVGEVKI 361  
 DB 303 DAKSVNELLHAGHTRHVKQISYIDSETEABGTALKNVDAILVYGGFGEVGEVKI 362  
 QY 362 STVRFARENKIPYLGICLQWQSAVIEFARNVYGLGASHTEFLPKSPHVIIGLITEMDE 421  
 DB 363 TIRAFARENKIPYLGICLQWQSAVIEFARNVYGLGASHTEFLPKSPHVIIGLITEMDE 422  
 QY 422 AGELVTRDESDLGCTMELGAKCKLADSLAFOLYXQDVITERRHRYEFERNQYLKOLE 481  
 DB 423 NAEVTRDESDLGCTMELGAKCKLADSLAFOLYXQDVITERRHRYEFERNQYLKOLE 482  
 QY 482 AAGKFSKSLDGRVLEIIELEPHEPFLACQFHEPFTSTPRNGHALFSGFVEAAK 537  
 DB 483 DAGLRVAGRSGDDQVLEIIELEPHEPFLACQFHEPFTSTPRNGHALFSGFVEAAK 538

RESULT 6  
 US-09-815-242-11175  
 Sequence 11175, Application US/09815242  
 Patent No. US20020061569A1  
 GENERAL INFORMATION:  
 APPLICANT: Haselbeck, Robert  
 APPLICANT: Ohlsen, Karl L.  
 APPLICANT: Zyskind, Judith W.  
 APPLICANT: Wall, Daniel  
 APPLICANT: Trawick, John D.  
 APPLICANT: Carr, Grant J.  
 APPLICANT: Yamamoto, Robert T.  
 APPLICANT: Xu, H. Howard  
 TITLE OF INVENTION: Identification of Essential Genes in  
 FILE REFERENCE: ELITRA.011A  
 CURRENT APPLICATION NUMBER: US/09/815,242  
 PRIOR FILING DATE: 2001-03-21  
 PRIOR APPLICATION NUMBER: 60/191,078  
 PRIOR FILING DATE: 2000-03-21  
 PRIOR APPLICATION NUMBER: 60/206,848  
 PRIOR FILING DATE: 2000-05-23  
 PRIOR APPLICATION NUMBER: 60/207,727  
 PRIOR FILING DATE: 2000-05-26

PRIOR APPLICATION NUMBER: 60/242,578  
 PRIOR FILING DATE: 2000-10-23  
 PRIOR APPLICATION NUMBER: 60/253,625  
 PRIOR FILING DATE: 2000-11-27  
 PRIOR APPLICATION NUMBER: 60/257,931  
 PRIOR FILING DATE: 2000-12-22  
 PRIOR APPLICATION NUMBER: 60/269,308  
 PRIOR FILING DATE: 2001-02-16  
 NUMBER OF SEQ ID NOS: 14110  
 SOFTWARE: FastSeq for windows version 4.0  
 SEQ ID NO 11175  
 LENGTH: 545  
 TYPE: PRT  
 ORGANISM: Haemophilus influenzae  
 US-09-815-242-11175

Query Match 68.5%; Score 1908.5; DB 9; Length 545;  
 Best Local Similarity 67.5%; Pred. No. 2.2e-171;  
 Matches 364; Conservative 71; Mismatches 103; Indels 1; Gaps 1;

QY 2 TKRFTITGVVSSLSGKGIASSLAALIEDRLKVTITLDPYINVDPGTMSPFQHGVEV 61  
 DB 3 TNYIFVTGVVSSLSGKGIASSLAALIEDRLKVTITLDPYINVDPGTMSPFQHGVEV 62  
 QY 62 TEDGATDLDLGHYERFLKTTMTKKNFTTQVYBOVLNERNKGDYLGATVQVPIHITDE 121  
 DB 63 TEDGATDLDLGHYERFLKTTMTKKNFTTQVYBOVLNERNKGDYLGATVQVPIHITDE 122  
 QY 122 IKRRVYESAEKDVALLIEVGTVGDIESLPLEFTIRQMGVELGRDRLAFHILTLVPIYKS 181  
 DB 123 IKRRVYESAEKDVALLIEVGTVGDIESLPLEFTIRQMGVELGRDRLAFHILTLVPIYKS 182  
 QY 182 AGELTKPTQHSVKELRTTIGIOPDILICRSEOPTPASBRKIALFTVAEKAVISAIDAD 241  
 DB 183 SEEVTKPTQHSVKELRTTIGIOPDILICRSDRAVPNERAKIALFCNVERAKVISLKYVD 242  
 QY 242 TIYRIPLLEKQGLDDLVVDQLRLDVPADLSAMEKVVYDGLTHPTDEVSIAIVKTYDHT 301  
 DB 243 SIYKIPGLKSGQLDDVYICRFSLNCPEANLSEWQVIFEEANPVSEVITIGWVKYIELP 302  
 QY 302 DAKSLNELLHAGHTRHVKQISYIDSETEABGTALKNVDAILVYGGFGEVGEVKI 361  
 DB 303 DAKSVNELLHAGHTRHVKQISYIDSETEABGTALKNVDAILVYGGFGEVGEVKI 362  
 QY 362 STVRFARENKIPYLGICLQWQSAVIEFARNVYGLGASHTEFLPKSPHVIIGLITEMDE 421  
 DB 363 TIRAFARENKIPYLGICLQWQSAVIEFARNVYGLGASHTEFLPKSPHVIIGLITEMDE 422  
 QY 422 AGELVTRDESDLGCTMELGAKCKLADSLAFOLYXQDVITERRHRYEFERNQYLKOLE 481  
 DB 423 NAEVTRDESDLGCTMELGAKCKLADSLAFOLYXQDVITERRHRYEFERNQYLKOLE 482  
 QY 482 AAGKFSKSLDGRVLEIIELEPHEPFLACQFHEPFTSTPRNGHALFSGFVEAAK-AKHX 539  
 DB 483 DAGLRVAGRSGDDQVLEIIELEPHEPFLACQFHEPFTSTPRNGHALFSGFVEAAK-AKHX 541

RESULT 7  
 US-09-815-242-13865  
 Sequence 13865, Application US/09815242  
 Patent No. US20020061569A1  
 GENERAL INFORMATION:  
 APPLICANT: Haselbeck, Robert  
 APPLICANT: Ohlsen, Karl L.  
 APPLICANT: Zyskind, Judith W.  
 APPLICANT: Wall, Daniel  
 APPLICANT: Trawick, John D.  
 APPLICANT: Carr, Grant J.  
 APPLICANT: Yamamoto, Robert T.  
 APPLICANT: Xu, H. Howard  
 TITLE OF INVENTION: Identification of Essential Genes in  
 FILE REFERENCE: ELITRA.011A

```

; CURRENT APPLICATION NUMBER: US/09/815,242
; CURRENT FILING DATE: 2001-03-21
; PRIOR APPLICATION NUMBER: 60/191,078
; PRIOR FILING DATE: 2000-03-21
; PRIOR APPLICATION NUMBER: 60/206,848
; PRIOR FILING DATE: 2000-05-23
; PRIOR APPLICATION NUMBER: 60/207,727
; PRIOR FILING DATE: 2000-05-26
; PRIOR APPLICATION NUMBER: 60/242,578
; PRIOR FILING DATE: 2000-10-23
; PRIOR APPLICATION NUMBER: 60/253,625
; PRIOR FILING DATE: 2000-11-27
; PRIOR APPLICATION NUMBER: 60/257,931
; PRIOR FILING DATE: 2000-12-22
; PRIOR APPLICATION NUMBER: 60/269,308
; PRIOR FILING DATE: 2001-02-16
; NUMBER OF SEQ ID NOS: 14110
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO: 13865
; LENGTH: 511
; TYPE: PRT
; ORGANISM: Salmonella typhi
; FEATURE:
; NAME/KEY: VARIANT
; LOCATION: (1)...(511)
; OTHER INFORMATION: Xaa = Any Amino Acid
US-09-815-242-13865

```

```

Query Match      64.0%; Score 1783; DB 9; Length 511;
Best Local Similarity 67.4%; Pred. No. 1,5e-159;
Matches 341; Conservative 64; Mismatches 101; Indels 0; Gaps 0;

```

```

QY 2 TKRIFITGVSSLSGKGIASSLAALIEBRLKVTITKDPYINVDGTMSPFOHGEVY 61
DB 3 TNYIFVTVGVSSLSGKGIASSLAALIEBRLKVTITKDPYINVDGTMSPFOHGEVY 62
QY 62 TEDGATDLDLGHYERFLKTKMTKKNFTTGOVYEOVLNBRKGDYLGATVOVPHITDE 121
DB 63 TEDGATDLDLGHYERFLKTKMTKKNFTTGOVYEOVLNBRKGDYLGATVOVPHITDE 122
QY 122 IKRVRVESABGVALLIEVGCTVGDIESLPLETTIQMGVEIGRPAALTIHTLPYIKS 181
DB 123 IKRVRVESABGVALLIEVGCTVGDIESLPLETTIQMGVEIGRPAALTIHTLPYIKS 182
QY 182 AGELETKRPOHSHKELRTIGIOPDILICRSEOPIPASERKIALFTNAEKAIVISAID 241
DB 183 AGELETKRPOHSHKELRTIGIOPDILICRSEOPIPASERKIALFTNAEKAIVISAID 242
QY 242 TIYRIFPLLEGGDLVDVQRLADVPAADLSAMEKRVVDGLTHPTDEVSIAIVGKYVDHT 301
DB 243 SIYKIFGLKSGQLDVIYICRFSIMCPKANTSEMEQVIYEANPAGRVITIGWGYKIEBP 302
QY 302 DAKSYLNEALIVGHTFRKVOISYIDSETIEAGTAKIKNDVAILVPGGFGFGRVGEKI 361
DB 303 DAKSYLNEALIVGHTFRKVOISYIDSETIEAGTAKIKNDVAILVPGGFGFGRVGEKI 362
QY 362 STYRFARENKIPYLGICLQKQSAVIEFARNVGLBGAHSTELPLKSPHPVIGLITEMWDE 421
DB 363 ATARVARENKIPYLGICLQKQSAVIEFARNVGLBGAHSTELPLKSPHPVIGLITEMWDE 422
QY 422 AGELVTRDESDIGTGMRLGAOKCRKADSLAFOLYQKOVITERRHRRYEFNNQYKOLE 481
DB 423 DGNVEVRSESDIGTGMRLGAOKCRKADSLAFOLYQKOVITERRHRRYEFNNQYKOLE 482
QY 482 AAGMKFSGKSLDGRLEIVEIIELEPHEPW 507
DB 483 AAGMKFSGKSLDGRLEIVEIIELEPHEPW 508

```

RESULT 8  
US-09-815-242-12372  
; Sequence 12372, Application US/09815242  
; Patent No. US20020061569A1

```

; GENERAL INFORMATION:
; APPLICANT: Haselbeck, Robert
; APPLICANT: Ohlsen, Kari L.
; APPLICANT: Zyskind, Judith W.
; APPLICANT: Wall, Daniel
; APPLICANT: Trawick, John D.
; APPLICANT: Carr, Grant J.
; APPLICANT: Yamamoto, Robert T.
; TITLE OF INVENTION: Identification of Essential Genes in
; FILE REFERENCE: ELITRA.011A
; CURRENT APPLICATION NUMBER: US/09/815,242
; CURRENT FILING DATE: 2001-03-21
; PRIOR APPLICATION NUMBER: 60/191,078
; PRIOR FILING DATE: 2000-03-21
; PRIOR APPLICATION NUMBER: 60/206,848
; PRIOR FILING DATE: 2000-05-23
; PRIOR APPLICATION NUMBER: 60/207,727
; PRIOR FILING DATE: 2000-05-26
; PRIOR APPLICATION NUMBER: 60/242,578
; PRIOR FILING DATE: 2000-10-23
; PRIOR APPLICATION NUMBER: 60/253,625
; PRIOR FILING DATE: 2000-11-27
; PRIOR APPLICATION NUMBER: 60/257,931
; PRIOR FILING DATE: 2000-12-22
; PRIOR APPLICATION NUMBER: 60/269,308
; PRIOR FILING DATE: 2001-02-16
; NUMBER OF SEQ ID NOS: 14110
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO: 12372
; LENGTH: 536
; TYPE: PRT
; ORGANISM: Staphylococcus aureus
US-09-815-242-12372

```

```

Query Match      56.5%; Score 1576; DB 9; Length 536;
Best Local Similarity 55.6%; Pred. No. 6e-140;
Matches 302; Conservative 93; Mismatches 134; Indels 14; Gaps 4;

```

```

QY 1 MKRFFITGVSSLSGKGIASSLAALIEBRLKVTITKDPYINVDGTMSPFOHGEVY 60
DB 1 MKRFFITGVSSLSGKGIASSLAALIEBRLKVTITKDPYINVDGTMSPFOHGEVY 60
QY 61 VTDEGATDLDLGHYERFLKTKMTKKNFTTGOVYEOVLNBRKGDYLGATVOVPHITD 120
DB 61 VTDEGATDLDLGHYERFLKTKMTKKNFTTGOVYEOVLNBRKGDYLGATVOVPHITD 120
QY 121 EIKRVRVESAE--GKVALIEVGCTVGDIESLPLETTIQMGVEIGRPAALTIHTLPY 178
DB 121 EIKRVRVESAE--GKVALIEVGCTVGDIESLPLETTIQMGVEIGRPAALTIHTLPY 180
QY 179 IKSAGELTKRPOHSHKELRTIGIOPDILICRSEOPIPASERKIALFTNAEKAIVISA 238
DB 181 IKSAGELTKRPOHSHKELRTIGIOPDILICRSEOPIPASERKIALFTNAEKAIVISA 240
QY 239 DADYTRIFPLLEGGDLVDVQRLADVPAADLSAMEKRVVDGLTHPTDEVSIAIVGKY 297
DB 241 DADYTRIFPLLEGGDLVDVQRLADVPAADLSAMEKRVVDGLTHPTDEVSIAIVGKY 300
QY 298 VDHTRVYSLSBALIHAGHTFRKVOISYIDSETIEAGTAKIKNDVAILVPGGFGFGRG 356
DB 301 VSLQDAIVSYVESLAKHAGTFRKVOISYIDSETIEAGTAKIKNDVAILVPGGFGFGRG 360
QY 357 VEGKISTVRFARENKIPYLGICLQKQSAVIEFARNVGLBGAHSTELPLKSPHPVIGLIT 416
DB 361 SEGKISAIKVRARENKIPYLGICLQKQSAVIEFARNVGLBGAHSTELPLKSPHPVIGLIT 420
QY 417 EEMDAGSLVTRDESDIGTGMRLGAOKCRKADSLAFOLYQKOVITERRHRRYEFNNQY 476
DB 421 EEMDAGSLVTRDESDIGTGMRLGAOKCRKADSLAFOLYQKOVITERRHRRYEFNNQY 470
QY 477 LKQLEMAKFKSGKSLDGRLEIVEIIELEPHEPWFLACQHPHEFTSTPRMGHALFSGFVEAAA 536

```

Db 471 RQLEANGVAVISGTSPPDRLEWVBIPTNDPFIACQFHEBFLSRPBRPHIFKSFIEASTL 530  
 QY 537 KHK 539  
 Db 531 KYQ 533

## RESULT 9

US-09-815-242-12978  
 ; Sequence 12978, Application US/09815242  
 ; Patent No. US20020061569A1  
 ; GENERAL INFORMATION:  
 ; APPLICANT: Haselbeck, Robert  
 ; APPLICANT: Ohlsen, Karl L.  
 ; APPLICANT: Zyskind, Judith W.  
 ; APPLICANT: Wall, Daniel  
 ; APPLICANT: Trawick, John D.  
 ; APPLICANT: Carr, Grant T.  
 ; APPLICANT: Yamamoto, Robert T.  
 ; APPLICANT: Xu, H. Howard  
 ; TITLE OF INVENTION: Identification of Essential Genes in  
 ; FILE REFERENCE: ELITRA.011A  
 ; CURRENT APPLICATION NUMBER: US/09/815,242  
 ; CURRENT FILING DATE: 2001-03-21  
 ; PRIOR APPLICATION NUMBER: 60/191,078  
 ; PRIOR FILING DATE: 2000-03-21  
 ; PRIOR APPLICATION NUMBER: 60/206,848  
 ; PRIOR FILING DATE: 2000-05-23  
 ; PRIOR APPLICATION NUMBER: 60/207,727  
 ; PRIOR FILING DATE: 2000-05-26  
 ; PRIOR APPLICATION NUMBER: 60/242,578  
 ; PRIOR FILING DATE: 2000-10-23  
 ; PRIOR APPLICATION NUMBER: 60/253,625  
 ; PRIOR FILING DATE: 2000-11-27  
 ; PRIOR APPLICATION NUMBER: 60/257,931  
 ; PRIOR FILING DATE: 2000-12-22  
 ; PRIOR APPLICATION NUMBER: 60/269,308  
 ; PRIOR FILING DATE: 2001-02-16  
 ; NUMBER OF SEQ ID NOS: 14110  
 ; SOFTWARE: FastSeq for Windows Version 4.0  
 ; SEQ ID NO 12978  
 ; LENGTH: 536  
 ; TYPE: PRT  
 ; ORGANISM: Staphylococcus aureus  
 ; US-09-815-242-12978

Query Match 56.5%; Score 1576; DB 9; Length 536;

Best Local Similarity 55.6%; Pred. No. 6e-140; Mismatches 134; Indels 14; Gaps 4;

Db 1 MKTKIFITGVVSSLSGKGIASSLAALIERGLKVTITKLDPIYINVDPGTMSFQGEV 60  
 1 MKTKIFITGVVSSLSGKGIASSLAALIERGLKVTITKLDPIYINVDPGTMSFQGEV 60  
 QY 61 VTEDEGATDLDLGHYERFLKTTMTKKNFTTGQVYEQVLRNKRKGDYLGATQVPIPHITD 120  
 61 VTEDEGATDLDLGHYERFLKTTMTKKNFTTGQVYEQVLRNKRKGDYLGATQVPIPHITD 120  
 Db 121 EIKRRVYESA--GKQVALLIEVGTVGDIESTLPLETIRQNGVELGRDALFIHLTLVPY 178  
 121 EIKRRVYESA--GKQVALLIEVGTVGDIESTLPLETIRQNGVELGRDALFIHLTLVPY 178  
 Db 121 EIKRRVYESA--GKQVALLIEVGTVGDIESTLPLETIRQNGVELGRDALFIHLTLVPY 180  
 121 EIKRRVYESA--GKQVALLIEVGTVGDIESTLPLETIRQNGVELGRDALFIHLTLVPY 180  
 QY 179 IKSAGELKTPQHSVLEKRTIGIOPDILICRSEOPIPASERRKIALFTVNAKAVISAI 238  
 179 IKSAGELKTPQHSVLEKRTIGIOPDILICRSEOPIPASERRKIALFTVNAKAVISAI 238  
 Db 181 IKAGEKKTPTQHSVLEKRTIGIOPDILICRSEOPIPASERRKIALFTVNAKAVISAI 240  
 181 IKAGEKKTPTQHSVLEKRTIGIOPDILICRSEOPIPASERRKIALFTVNAKAVISAI 240  
 QY 239 DADITVPIPLKEQGLDVLVNDLRDVP--AADLSMEKVDGLTFPDBVSIATV 297  
 239 DADITVPIPLKEQGLDVLVNDLRDVP--AADLSMEKVDGLTFPDBVSIATV 297  
 Db 241 DADSLYEIPQLQSQNDVDIVIKRLQINAKYETQLDWMKQLDIIYNNLDKRTITIGLVGKI 300  
 241 DADSLYEIPQLQSQNDVDIVIKRLQINAKYETQLDWMKQLDIIYNNLDKRTITIGLVGKI 300  
 QY 298 VDHTEAVKSLNEALIHGSIHTRHKRVQISYIDSEETIEAGTAK--LKNVDAILVPGGREG 356  
 298 VDHTEAVKSLNEALIHGSIHTRHKRVQISYIDSEETIEAGTAK--LKNVDAILVPGGREG 356

Db 301 VSLQDAVLSVESLKHGYPPAKDIDIRWIDSSVTDENAEVLAADVGLIVPGGPRRA 360  
 QY 357 VEGKISTVRFARENKIFPIYLIGICMGQSAVIEFANVVGLEGAHSTBELPKSPHFVIGIT 416  
 357 VEGKISTVRFARENKIFPIYLIGICMGQSAVIEFANVVGLEGAHSTBELPKSPHFVIGIT 416  
 Db 361 SEGKISAKTVARNENNVFFPGICMGQLATVFPSSNVVGLBGAHSELDPAFPYPIIDLP 420  
 361 SEGKISAKTVARNENNVFFPGICMGQLATVFPSSNVVGLBGAHSELDPAFPYPIIDLP 420  
 QY 417 EMDPAGELVTRDSDSLGTMRLGAQCKRLKADSLAFOLYQKQVITERHRRYEFNNQY 476  
 417 EMDPAGELVTRDSDSLGTMRLGAQCKRLKADSLAFOLYQKQVITERHRRYEFNNQY 476  
 Db 421 EQKD-----IEDLGGTIRLGLYPCSIXEGTLADQVYGAEIERHRRYEFNNQY 470  
 421 EQKD-----IEDLGGTIRLGLYPCSIXEGTLADQVYGAEIERHRRYEFNNQY 470  
 QY 477 LKOLRAAGMFGSKSLQGRVLEIIELEPHWFLACQFHEBFLSRPBRPHIFKSFIEASTL 536  
 477 LKOLRAAGMFGSKSLQGRVLEIIELEPHWFLACQFHEBFLSRPBRPHIFKSFIEASTL 536  
 Db 471 RQLEANGVAVISGTSPPDRLEWVBIPTNDPFIACQFHEBFLSRPBRPHIFKSFIEASTL 530  
 471 RQLEANGVAVISGTSPPDRLEWVBIPTNDPFIACQFHEBFLSRPBRPHIFKSFIEASTL 530  
 QY 537 KHK 539  
 Db 531 KYQ 533

## RESULT 10

US-09-815-242-4984  
 ; Sequence 4984, Application US/09815242  
 ; Patent No. US20020061569A1  
 ; GENERAL INFORMATION:  
 ; APPLICANT: Haselbeck, Robert  
 ; APPLICANT: Ohlsen, Karl L.  
 ; APPLICANT: Zyskind, Judith W.  
 ; APPLICANT: Wall, Daniel  
 ; APPLICANT: Trawick, John D.  
 ; APPLICANT: Carr, Grant T.  
 ; APPLICANT: Yamamoto, Robert T.  
 ; APPLICANT: Xu, H. Howard  
 ; TITLE OF INVENTION: Identification of Essential Genes in  
 ; FILE REFERENCE: ELITRA.011A  
 ; CURRENT APPLICATION NUMBER: US/09/815,242  
 ; CURRENT FILING DATE: 2001-03-21  
 ; PRIOR APPLICATION NUMBER: 60/191,078  
 ; PRIOR FILING DATE: 2000-03-21  
 ; PRIOR APPLICATION NUMBER: 60/206,848  
 ; PRIOR FILING DATE: 2000-05-23  
 ; PRIOR APPLICATION NUMBER: 60/207,727  
 ; PRIOR FILING DATE: 2000-05-26  
 ; PRIOR APPLICATION NUMBER: 60/242,578  
 ; PRIOR FILING DATE: 2000-10-23  
 ; PRIOR APPLICATION NUMBER: 60/253,625  
 ; PRIOR FILING DATE: 2000-11-27  
 ; PRIOR APPLICATION NUMBER: 60/257,931  
 ; PRIOR FILING DATE: 2000-12-22  
 ; PRIOR APPLICATION NUMBER: 60/269,308  
 ; PRIOR FILING DATE: 2001-02-16  
 ; NUMBER OF SEQ ID NOS: 14110  
 ; SOFTWARE: FastSeq for Windows Version 4.0  
 ; SEQ ID NO 4984  
 ; LENGTH: 536  
 ; TYPE: PRT  
 ; ORGANISM: Enterococcus faecalis  
 ; US-09-815-242-4984

Query Match 56.3%; Score 1569; DB 9; Length 536;

Best Local Similarity 55.4%; Pred. No. 2.8e-139; Mismatches 134; Indels 14; Gaps 4;

Db 1 MKTKIFITGVVSSLSGKGIASSLAALIERGLKVTITKLDPIYINVDPGTMSFQGEV 60  
 1 MKTKIFITGVVSSLSGKGIASSLAALIERGLKVTITKLDPIYINVDPGTMSFQGEV 60  
 QY 61 VTEDEGATDLDLGHYERFLKTTMTKKNFTTGQVYEQVLRNKRKGDYLGATQVPIPHITD 120  
 61 VTEDEGATDLDLGHYERFLKTTMTKKNFTTGQVYEQVLRNKRKGDYLGATQVPIPHITD 120  
 Db 61 VTEDEGATDLDLGHYERFLKTTMTKKNFTTGQVYEQVLRNKRKGDYLGATQVPIPHITD 120  
 61 VTEDEGATDLDLGHYERFLKTTMTKKNFTTGQVYEQVLRNKRKGDYLGATQVPIPHITD 120  
 QY 121 EIKRRVYESA--GKQVALLIEVGTVGDIESTLPLETIRQNGVELGRDALFIHLTLVPY 178  
 121 EIKRRVYESA--GKQVALLIEVGTVGDIESTLPLETIRQNGVELGRDALFIHLTLVPY 178

Db	121	ETREKMPAAKTTDADVITTEVGIVGDIESPLELEALRQKAKMGSDNNYIHTTILIF	180
Qy	179	IKSAGELTKPTQHSVKELRTIGIDPILICRSEOPIPASERKIALFTVAEKAVISAL	238
Db	181	IKAGSMKTKPTQHSVKELRTSIGIDPILIVRTTELTPVSGNTKTKLAQCDVNEPAVIESR	240
Qy	239	DADTYIRIPELLREBGLDLDVDDLDLDVPAADUSAMERYVDGLTHPDEVISALVKRY	298
Db	241	DVETLTSIPILAQAMQDQIVCDHKLDPADMTENALEEKYLNKKKTKJALVKRY	300
Qy	299	DHTDAKSLNELIHAHGITHREKQDISYIDSEIIBAGTA-KLMVDAILVPGSGERG	357
Db	301	ELPDVAISVEALEAKAGFDPSDIELDWDVDSQELTAENVARISGADGILVPGSGRGI	360
Qy	358	EGRISVTRPAREKQPYLGICGAMOSAVIEPARNVGLGASHTEFLPKSPHYIGITE	417
Db	361	EGRILEIRPARENDVDFGICGQMOMACEPERNNVGLDGSGAETNDDVTNNIIDLMA	420
Qy	418	WMDEAGELVTRDESDSLGSTMELGAKCKMLKADSLAFOLY-QDYVTERHRYERNQY	476
Db	421	-----QENIEMLGTLRLGLVPCRLKKGFTAAAYAGNEDVQERHHRHREPNKY	470
Qy	477	LKOLEAAGKFSKSLDGRVLELLELPHRPWELAQEPETSPRNQHALFSGFVAAA	536
Db	471	ROFEEENGLVFSVSPDNRLVEIVEIPEKQFVAAQDFPHELISRNPRORLIKQFVGAL	530
Qy	537	KHK 539	
Db	531	ANK 533	

```

RESULT 11
US-09-815-242-10727
: Sequence 10727, Application US/09815242
: Patent No. US20020061569a1
: GENERAL INFORMATION:
: APPLICANT: Haselbeck, Robert
: APPLICANT: Ohlsen, Karl L.
: APPLICANT: Zyekind, Judith W.
: APPLICANT: Wall, Daniel
: APPLICANT: Trawick, John D.
: APPLICANT: Carr, Grant J.
: APPLICANT: Yamamoto, Robert T.
: APPLICANT: Xu, H. Howard
: TITLE OF INVENTION: Identification of Essential Genes in
: TITLE OR INVENTION: Prokaryotes
: FILE REFERENCE: EUTRA.011a
: CURRENT APPLICATION NUMBER: US/09/815,242
: PRIOR FILING DATE: 2001-03-21
: PRIOR APPLICATION NUMBER: 60/191,078
: PRIOR FILING DATE: 2000-03-21
: PRIOR APPLICATION NUMBER: 60/206,848
: PRIOR FILING DATE: 2000-05-23
: PRIOR APPLICATION NUMBER: 60/207,727
: PRIOR FILING DATE: 2000-05-26
: PRIOR APPLICATION NUMBER: 60/242,578
: PRIOR FILING DATE: 2000-10-23
: PRIOR APPLICATION NUMBER: 60/253,625
: PRIOR FILING DATE: 2000-11-27
: PRIOR APPLICATION NUMBER: 60/257,931
: PRIOR FILING DATE: 2000-12-22
: PRIOR APPLICATION NUMBER: 60/269,308
: PRIOR FILING DATE: 2001-02-16
: NUMBER OF SEQ ID NOS: 14110
: SOFTWARE: FastSeq for Windows Version 4.0
: SEQ ID NO 10727
: LENGTH: 536
: TYPE: PRT
: ORGANISM: Enterococcus faecalis
US-09-815-242-10727
Query Match      56.3%, Score 1569, DB 9, Length 536

```

[illegible]

```

1 RESULT 12
2
3 ; Sequence 13376, Application US/09815242
4 ; Patent No. US20020061569A1
5
6 GENERAL INFORMATION:
7
8 APPLICANT: Haselbeck, Robert
9 APPLICANT: Ohlsem, Kari L.
10 APPLICANT: Zvekind, Judith W.
11 APPLICANT: Wall, Daniel
12 APPLICANT: Trewick, John D.
13 APPLICANT: Carr, Grant J.
14 APPLICANT: Yamamoto, Robert T.
15 APPLICANT: Xu, H. Howard
16
17 TITLE OF INVENTION: Identification of Essential Genes in
18 TITLE OR INVENTION: Prokaryotes
19
20 TITLE REFERENCE: ELITRA.0114
21
22 CURRENT APPLICATION NUMBER: US/09/815,242
23
24 PRIOR FILING DATE: 2001-03-21
25
26 PRIOR APPLICATION NUMBER: 60/139,171
27
28 PRIOR FILING DATE: 2000-03-21
29
30 PRIOR APPLICATION NUMBER: 60/206,848
31
32 PRIOR FILING DATE: 2000-05-23
33
34 PRIOR APPLICATION NUMBER: 60/207,727
35
36 PRIOR FILING DATE: 2000-05-26
37
38 PRIOR APPLICATION NUMBER: 60/242,578
39
40 PRIOR FILING DATE: 2000-10-23
41
42 PRIOR APPLICATION NUMBER: 60/253,625
43
44 PRIOR FILING DATE: 2000-11-27
45
46 PRIOR APPLICATION NUMBER: 60/257,931

```

PRIOR FILING DATE: 2000-12-22  
 PRIOR APPLICATION NUMBER: 60/269,308  
 PRIOR FILING DATE: 2001-02-16  
 NUMBER OF SEQ ID NOS: 14110  
 SOFTWARE: FastSeq for Windows Version 4.0  
 SEQ ID NO: 13376  
 LENGTH: 535  
 TYPE: PRT  
 ORGANISM: Streptococcus pneumoniae  
 US-09-815-242-13376

Query Match 56.0%; Score 1561; DB 9; Length 535;  
 Best Local Similarity 54.3%; Pred. No. 1,66-138;  
 Matches 294; Conservative 108; Mismatches 125; Indels 14; Gaps 4;

2 TKFIFITGVSSLSGKGAASLAILEDRLKVTITKLDPIYINVDPGTMSPFQHGVEVY 61  
 3 TKYIFVTGVSSLSGKGAASLAILEDRLKVTITKLDPIYINVDPGTMSPFQHGVEVY 62  
 62 TDDGATDLDLGHYRFRFLTKMTKKNFTQVYBQVLRNKRKGYLGATVQVPHITDE 121  
 63 TDDGATDLDLGHYRFRFLTKMTKKNFTQVYBQVLRNKRKGYLGATVQVPHITDA 122  
 122 IKRRYESA--EGKVALTEVGTVGDIESTLPLETTRQWVELGRBALFHTLVPIYI 179  
 123 LKFKIKRALTTDSVITTEVGTVGDIESTLPLETTRQWVELGRBALFHTLVPIYI 182  
 180 KSAGSLKPTQHSYKELRTIGIOPDILCRSEOPTPASERKIALFTNVAEKAVISAD 239  
 183 KKAAGMKKPTQHSYKELRTIGIOPDILCRSEOPTPASERKIALFTNVAEKAVISAD 242  
 240 ADTIRIPFLLRBQGLDLDVVDQLRDVPAADLSAMKKNVVDGLTPTDEVSIAIVGYVD 299  
 243 VEHVQIPLNLAQGMDOVCHLKDAPADMTMSAMVDKVMKLVKKQVMSLVGKYVE 302  
 300 HTDAVKSLEALIHAGIHTRHVKQISYIDSETTEAEGLNK--LKNVDALIVPGFGEKVE 358  
 303 LQDAVISVEALKHSGVNDVVKINWVANNDVTAEVALLSDADGIIVPGFGQRGTE 362  
 359 GKTSTWFRKRNKIPYIGICLGMQSAVIEFANVVGLEGAHSTELPKSPHPIVGLIEM 418  
 363 GKIQIRKARENDVPLSGVCLGMQSLCIEFAHVTGLBGANASALAPETKVIYIIMDQ 422  
 419 MDAGELVTRDESDLGTMRLGAQKCRILKADSLAFQLY--QKVITERRHRYEENNOYL 477  
 423 ID-----IEDWGTIRLGLVYSKLRGSKKAAAHVNOEVRQRRHRYEENNAFR 472  
 478 KQLEAGMKPFGSKSLDGRVLEIIELEHFWPLACQHPFETSTPNNGHALFSGFVEAAK 537  
 473 EGFPAAGFVFSVPNRLVLEIIELEHFWPLACQHPFETSTPNNGHALFSGFVEAAK 532  
 QY 538 H 538  
 DB 533 N 533

RESULT 13  
 US-09-815-242-11320  
 Sequence 11320, Application US/09815242  
 Patent No. US20020061569A1  
 GENERAL INFORMATION:  
 APPLICANT: Haselbeck, Robert  
 APPLICANT: Ohlsen, Kati L.  
 APPLICANT: Zyskind, Judith W.  
 APPLICANT: Wall, Daniel  
 APPLICANT: Trawick, John D.  
 APPLICANT: Carr, Grant J.  
 APPLICANT: Yamamoto, Robert T.  
 APPLICANT: Xu, H. Howard  
 TITLE OF INVENTION: Identification of Essential Genes in  
 FILE REFERENCE: ELITRA, 011A  
 CURRENT APPLICATION NUMBER: US/09/815, 242

CURRENT FILING DATE: 2001-03-21  
 PRIOR APPLICATION NUMBER: 60/191,078  
 PRIOR FILING DATE: 2000-03-21  
 PRIOR APPLICATION NUMBER: 60/206,848  
 PRIOR FILING DATE: 2000-05-23  
 PRIOR APPLICATION NUMBER: 60/207,727  
 PRIOR FILING DATE: 2000-05-26  
 PRIOR APPLICATION NUMBER: 60/242,578  
 PRIOR FILING DATE: 2000-10-23  
 PRIOR APPLICATION NUMBER: 60/253,625  
 PRIOR FILING DATE: 2000-11-27  
 PRIOR APPLICATION NUMBER: 60/257,931  
 PRIOR FILING DATE: 2000-12-22  
 PRIOR APPLICATION NUMBER: 60/269,308  
 PRIOR FILING DATE: 2001-02-16  
 NUMBER OF SEQ ID NOS: 14110  
 SOFTWARE: FastSeq for Windows Version 4.0  
 SEQ ID NO: 11320  
 LENGTH: 538  
 TYPE: PRT  
 ORGANISM: Helicobacter pylori  
 US-09-815-242-11320

Query Match 51.8%; Score 1444.5; DB 9; Length 538;  
 Best Local Similarity 50.8%; Pred. No. 1,76-127;  
 Matches 271; Conservative 112; Mismatches 147; Indels 3; Gaps 2;

3 KPIFTGVSSLSGKGAASLAILEDRLKVTITKLDPIYINVDPGTMSPFQHGVEVY 62  
 5 KPIFTGVSSLSGKGAASLAILEDRLKVTITKLDPIYINVDPGTMSPFQHGVEVY 64  
 63 EDGATDLDLGHYRFRFLTKMTKKNFTQVYBQVLRNKRKGYLGATVQVPHITDEI 122  
 65 SDGATDLDLGHYRFRFLTKMTKKNFTQVYBQVLRNKRKGYLGATVQVPHITDEI 124  
 123 KRRYESAEGDVALTEVGTVGDIESTLPLETTRQWVELGRBALFHTLVPIYKSA 182  
 125 KRRKSAKAGLDPLTEVGTVGDIESTLPLETTRQWVELGRBALFHTLVPIYKSA 184  
 183 GELTKPTQHSYKELRTIGIOPDILCRSEOPTPASERKIALFTNVAEKAVISADOT 242  
 185 NELTKPTQHSYKELRTIGIOPDILCRSEOPTPASERKIALFTNVAEKAVISADOT 244  
 243 IYRIPFLLRBQGLDLDVVDQLRDVPAADLSAMKKNVVDGLTPTDEVSIAIVGYVHTD 302  
 245 IYAPPIIFLQBGILTPARFENLAKLHPKAAAMNTLVKTIAPKAKYKLVGKYLISKE 304  
 303 AYKSLNEALIHAGIHTRHVKQISYIDSETTEAEGLNK--LKNVDALIVPGFGEKVE 362  
 305 SYKSLIEALIHAGIHTRHVKQISYIDSETTEAEGLNK--LKNVDALIVPGFGEKVE 362  
 363 TWPARENKIPYIGICLGMQSAVIEFANVVGLEGAHSTELPKSPHPIVGLIEMWDEA 422  
 363 AIQARLEKLPYIGICLGMQSAVIEFANVVGLEGAHSTELPKSPHPIVGLIEMWDEA 422  
 423 GELVTRDESDLGTMRLGAQKCRILKADSLAFQLYQKVITERRHRYEENNOYLKQLEA 482  
 423 HQQVRTYNSPLSGTMRLGERCEIMNSLLEKAYKPSIKERRHRYEENNOYLKQLEA 482  
 483 AKMFPKSKSLDGRVLEIIELEHFWPLACQHPFETSTPNNGHALFSGFVEAA 535  
 DB 483 KGLKVVGFG--SNHLIEALEHDPFVGVQFHPETSTPNNGHALFSGFVEAA 534

RESULT 14  
 US-10-156-761-14031  
 Sequence 14031, Application US/10156761  
 Publication No. US20030119018A1  
 GENERAL INFORMATION:  
 APPLICANT: OMURA, SATOSHI  
 APPLICANT: IKEDA, HARUO  
 APPLICANT: ISHIKAWA, JUN  
 APPLICANT: HORIKAWA, HIROSHI

```

; APPLICANT: SHIBA, TADAYOSHI
; APPLICANT: SAKAKI, YOSHITAKI
; APPLICANT: HATTORI, MASAHIRA
; TITLE OF INVENTION: NOVEL POLYNUCLEOTIDES
; FILE REFERENCE: 249-262
; CURRENT APPLICATION NUMBER: US/10/156,761
; CURRENT FILING DATE: 2002-05-29
; PRIOR APPLICATION NUMBER: JP 2001-204089
; PRIOR FILING DATE: 2001-05-30
; PRIOR APPLICATION NUMBER: JP 2001-272697
; PRIOR FILING DATE: 2001-08-02
; NUMBER OF SEQ ID NOS: 15109
; SEQ ID NO 14031
; LENGTH: 549
; TYPE: PRT
; ORGANISM: Streptomyces avermitilis
; US-10-156-761-14031

```

```

Query Match      51.4%; Score 1432.5; DB 15; Length 549;
Best Local Similarity 52.5%; Pred. No. 2.3e-126;
Matches 287; Conservative 92; Mismatches 153; Indels 15; Gaps 7;

```

```

QY 2 TKFIFITGCVSSLSKGIASLSLAILEDRLKVTITKLDPIYINVDPGTMSPPQGEVVF 61
   |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
DB 8 TKHIFVTGCVSSLSKGIASLSLAILEDRLKVTITKLDPIYINVDPGTMSPPQGEVVF 67
   |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
QY 62 TEDGATDLDLGHYERFLKTTMKKNFTTGQVYEOVLNERKGDYLGATVQVIPHIDE 121
   |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
DB 68 TNDGATDLDLGHYERFLKTTMKKNFTTGQVYEOVLNERKGDYLGATVQVIPHIDE 127
   |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
QY 122 IKRKYVESAEK-DVALIEVGTVGDIESLPLETRIRGMSVGLGRDALFIHLTVPIYK 180
   |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
DB 128 IKRIRIRKATDEVDAVITEVGTVGDIESLPLETRIRGMSVGLGRDALFIHLTVPIYK 187
   |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
QY 181 SAGELTKEPTQSVKELRTIGIOPDILICRSEQPIASERKIALFTNVAEKAVISAIDA 240
   |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
DB 188 PEGELKTKPTQSVKELRTIGIOPDILICRSEQPIASERKIALFTNVAEKAVISAIDA 247
   |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
QY 241 DTIYRIPLEIRREGDLDLVVDOLRLDVPADLSAMEKVVGDLTPTDEVSIAIVGYVDH 300
   |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
DB 248 RSTIYDPEKTVHGGDADAYVRKIDLPFRVDWTMTDLDLRHAKPDHEINLAVGKIYDL 307
   |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
QY 301 TDVYKSLNRLIHAGIHTHRKQVISTIDSETIE--ABGTAKLKNVDAILVPGGFERGYE 358
   |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
DB 308 PAVYISVTALAGGAPNARARVAKIKWTSDDCTPAGAGQGLDVAIDICPGAFGDRGVS 367
   |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
QY 359 GKISTVPARENKIPYIGICIGMSAVIEFARNVGLLEGASTEPFSPKSPHPIGLITEM 418
   |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
DB 368 GKVGAIQVAREHRIPLIGLGLQCTVIEAARNLADIPRANSTBPSATGHEFVISTMAEQ 427
   |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
QY 419 MD-EAGELVTRDESDLGTMRLGAQCKRLKADSLAPOLYO-KDVTTERHRRHYEPNNQY 476
   |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
DB 428 LDIIVAGS-----GDMGTMRLGMPALAECSIVREYDGEYVEEHRHRYEYNNAY 480
   |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
QY 477 LWOLE-AAGCKFSKSLDGRIVEIIELEP--HPWFLACQFHPFSTPRNGALSSGPVE 533
   |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
DB 481 RAELEKXAGLQSGTSBDGLVEYVPRVHPYIVATQAHPELRSRPTRPPLPAGLVK 540
   |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
QY 534 AAAXKXT 540
   |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
DB 541 AAVERTK 547
   |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:

```

```

RESULT 15
US-09-738-626-5066
; Sequence 5066, Application US/09738626
; Publication No. US20020197605A1
; GENERAL INFORMATION:
; APPLICANT: NAKAGAWA, SATOSHI
; APPLICANT: MIZOGUCHI, HIROSHI
; APPLICANT: ANDO, SEIKO
; APPLICANT: HAYASHI, MIKIRO
; APPLICANT: OCHIAI, KEIKO

```

```

; APPLICANT: YOKOI, HARUHIKO
; APPLICANT: YATEISHI, NAOKO
; APPLICANT: SENDO, AKIHIRO
; APPLICANT: IKEDA, MASATO
; APPLICANT: OZAKI, AKIO
; TITLE OF INVENTION: NOVEL POLYNUCLEOTIDES
; FILE REFERENCE: 249-125
; CURRENT APPLICATION NUMBER: US/09/738,626
; CURRENT FILING DATE: 2000-12-18
; PRIOR APPLICATION NUMBER: JP 99/377484
; PRIOR FILING DATE: 1999-12-16
; PRIOR APPLICATION NUMBER: JP 00/159162
; PRIOR FILING DATE: 2000-04-07
; PRIOR APPLICATION NUMBER: JP 00/280988
; PRIOR FILING DATE: 2000-08-03
; NUMBER OF SEQ ID NOS: 7059
; SOFTWARE: Patent in ver. 3.0
; SEQ ID NO 5066
; LENGTH: 554
; TYPE: PRT
; ORGANISM: Corynebacterium glutamicum
; US-09-738-626-5066

```

```

Query Match      49.2%; Score 1372; DB 10; Length 554;
Best Local Similarity 51.9%; Pred. No. 1.3e-120;
Matches 283; Conservative 78; Mismatches 166; Indels 18; Gaps 6;

```

```

QY 2 TKFIFITGCVSSLSKGIASLSLAILEDRLKVTITKLDPIYINVDPGTMSPPQGEVVF 61
   |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
DB 10 TKHIFVTGCVSSLSKGIASLSLAILEDRLKVTITKLDPIYINVDPGTMSPPQGEVVF 69
   |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
QY 62 TEDGATDLDLGHYERFLKTTMKKNFTTGQVYEOVLNERKGDYLGATVQVIPHIDE 121
   |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
DB 70 TEDGATDLDLGHYERFLKTTMKKNFTTGQVYEOVLNERKGDYLGATVQVIPHIDE 129
   |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
QY 122 IKRKYVESAE-----GKOVALLRVGTVGDIESLPLETRIRGMSVGLGRDALFIHLTVL 175
   |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
DB 130 IKARILSMGEEDPAGNADVAIVSEVGTVGDIESQPLEAARQVYHIEGRENCFPIHCSL 189
   |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
QY 176 VPYKSAEELKTKPTQSVKELRTIGIOPDILICRSEQPIASERKIALFTNVAEKAVI 235
   |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
DB 190 VPIYATSELSKTKPTQSVKELRTIGIOPDILICRSEQPIASERKIALFTNVAEKAVI 249
   |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
QY 236 SAIDADTIYRIPLEIRREGDLDLVVDOLRLDVPADLSAMEKVVGDLTPTDEVSIAIVG 295
   |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
DB 250 SCPOSSSIYINIPDYLREHLLPTFIIRLGLPFRVDWSTWHDLERVNNPHELTGIVG 309
   |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
QY 296 KYVDHTDAVYKSLNRLIHAGIHTHRKQVISTIDSETIEAGTA--KLKNVDAILVPGGFG 353
   |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
DB 310 KYIDLPDAIVSVAVRAAGYANNTRINIKWTSDDCTPAGAKKALSGLDVAIVPGGFG 369
   |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
QY 354 BRGVGKISTVPARENKIPYIGICIGMSAVIEFARNVGLLEGASTEPFSPKSPHPIYIG 413
   |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
DB 370 IRGIEGKIGALTFRREHKIPILIGLGLQCTVIEAARNLADIPRANSTBPSATGHEFVISTMAEQ 428
   |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
QY 414 LITWMDAAGELVTRDESDLGTMRLGAQCKRLKADSLAPOLYO-KDVTTERHRRHYEPN 473
   |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
DB 429 TMBE-----OKAAVSGADLGGTMRLGAYPATIEESLVEIYOTTEVSESRHRRHYVN 482
   |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
QY 474 NOYLKOL-EAAGCKFSKSLDGRIVEIIELEP--HPWFLACQFHPFSTPRNGALSSGPVE 530
   |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
DB 483 NAYRAQIEGSDLVFSGTSPGHIVERVEYKAEHPYIVATQAHPELRSRPTRPPLPAGLVK 542
   |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:

```

```

Search completed: January 29, 2004, 16:21:18
Job time : 38.1346 secs

```

GenCore version 5.1.6  
Copyright (c) 1993 - 2004 CompuGen Inc.

OM protein - protein search, using sw model

Run on: January 29, 2004, 15:44:49 ; Search time 16.0063 Seconds  
(without alignments)  
3268.453 Million cell updates/sec

Title: US-09-941-947a-16

Perfect score: 2788  
Sequence: 1 MTKRFFITGGVSSLGKGLA.....HALPSGFVEAAAKHTQCTA 544

Scoring table: BLOSUM62  
Gapop 10.0 , Gapext 0.5

Searched: 283308 segs, 96168682 residues

Total number of hits satisfying chosen parameters: 283308

Minimum DB seq length: 0  
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Listing first 45 summaries

Database :  
1: pir1:\*  
2: pir2:\*  
3: pir3:\*  
4: pir4:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

## SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	2008	72.0	542	B83192	CTP synthase PA363
2	1946	69.8	545	AD0859	CTP synthetase [im
3	1934	69.0	545	E82074	CTP synthase VC244
4	1916	68.7	545	1 STECTP	CTP synthase (EC 6
5	1916	68.7	545	H91083	CTP synthetase [im
6	1916	68.7	545	A85929	CTP synthetase [im
7	1909	68.5	545	AC0410	CTP synthase (EC 6
8	1908.5	68.5	545	PC4181	CTP synthase (EC 6
9	1905	68.3	554	C82700	CTP synthetase XPI
10	1892.5	67.9	544	PC1798	CTP synthase (EC 6
11	1888.5	67.7	544	PC1070	CTP synthase NMB15
12	1737	62.3	545	B84978	CTP synthase (EC 6
13	1645.5	59.0	532	H84123	CTP synthetase ctr
14	1639	58.8	531	P70415	CTP synthetase - A
15	1622.5	57.9	535	1 SYRSTP	CTP synthase (EC 6
16	1615.5	57.9	535	C97255	CTP synthase (UTP-
17	1603.5	57.5	552	CT75840	CTP synthase (EC 6
18	1600.5	57.4	532	AB1770	CTP synthases homo
19	1600.5	57.4	532	AG1394	CTP synthases homo
20	1595	57.2	550	D87462	CTP synthase [impo
21	1592	57.1	542	AC3358	CTP synthase (EC 6
22	1576	56.5	545	D90006	CTP synthetase [impo
23	1574.5	56.5	545	AH2430	CTP synthetase [impo
24	1563.5	56.1	540	E64446	CTP synthase [impo
25	1562	56.0	535	C95057	CTP synthase [impo
26	1562	56.0	578	AF2775	CTP synthase [impo
27	1562	56.0	578	D97555	CTP synthase (UTP-
28	1561	55.8	535	P79926	CTP synthase (EC 6
29	1555.5	55.8	537	P75149	CTP synthase (pyrG

30	1542	55.3	544	2	S25101	CTP synthase (EC 6
31	1538.5	55.2	557	2	H71189	probable CTP synth
32	1507	54.1	543	2	E81418	CTP synthase (EC 6
33	1484	53.2	533	2	E69154	CTP synthetase - Met
34	1449	52.0	538	2	C71946	CTP synthetase - Hel
35	1444.5	51.8	538	2	E64563	CTP synthetase - H
36	1441.5	51.7	549	2	C76879	probable CTP synth
37	1434	51.4	535	2	C86686	CTP synthetase [im
38	1423	51.0	544	2	B75379	CTP synthase - Del
39	1374.5	49.3	532	2	D69281	CTP synthase (pyrG
40	1369.5	49.1	577	2	C71342	probable CTP synth
41	1365.5	49.0	586	2	B70503	probable pyrG prot
42	1362	48.9	553	2	B84334	CTP synthase [impo
43	1356.5	48.7	537	2	A72103	CTP synthetase - C
44	1356	48.6	535	2	D90161	CTP synthetase (UT
45	1355.5	48.6	537	2	D86520	CTP synthetase [im

## ALIGNMENTS

## RESULT 1

B83192 CTP synthase PA3637 [imported] - Pseudomonas aeruginosa (strain PA01)

C:Species: Pseudomonas aeruginosa

C:Date: 15-Sep-2000 #sequence\_revision 15-Sep-2000 #text\_change 31-Dec-2000

C:Accession: B83192

R:Stover, C.K.; Pham, X.O.; Erwin, A.L.; Mizoguchi, S.D.; Warener, P.; Hickey, M.J.; Bri

adman, S.; Yuan, Y.; Brody, L.L.; Coulter, S.N.; Folger, K.R.; Kas, A.; Laidig, K.; Lim,

.; Loiry, S.; Olson, M.V.

Nature 406, 959-964, 2000

A:Title: Complete genome sequence of Pseudomonas aeruginosa PA01, an opportunistic pathog

A:Reference number: A82950; MUID:20437337; PMID:10984043

A:Accession: B83192

A:Status: preliminary

A:Molecule type: DNA

A:Residues: 1-542 <STD>

A:Cross-references: GB:AE004783; GB:AE004091; NID:99949786; PIDN:AA07025.1; GSPDB:GNO01;

A:Experimental source: strain PA01

C:Genetics:

A:Gene: pyrG; PA3637

C:Superfamily: CTP synthase

Query Match 72.0%; Score 2008; DB 2; Length 542;

Best Local Similarity 71.2%; Pred. No. 3.1e-121;

Matches 383; Conservative 64; Mismatches 91; Indels 0; Gaps 0;

QY	1	MTKRFFITGGVSSLGKGLAASSIAATLEDRGLKVTITKDPYINVPQTSPFGHGEVF	60
DB	1	MTKRFFITGGVSSLGKGLASASIAATLEDRGLKVTITKDPYINVPQTSPFGHGEVF	60
QY	61	VTBGAFTDLDLGHYERFLKTKTKKNNFTTGQYEVOLRNERKQDYLGAIVQVIPHITD	120
DB	61	VTBGAFTDLDLGHYERFLKTKTKKNNFTTGQYEVOLRNERKQDYLGAIVQVIPHITD	120
QY	121	EIKRKYVESAGKQVALIEVGCTVGDIESLPFLETTIQMGYELGRDALFHTLTPPIK	180
DB	121	EIKRKYVESAGKQVALIEVGCTVGDIESLPFLETTIQMGYELGRDALFHTLTPPIK	180
QY	181	SAGEIKTPQHSVKEKRTIGIOPDILICRSEOPTPASERKIALFTVNAKAVISAIDA	240
DB	181	SAGEIKTPQHSVKEKRTIGIOPDILICRSEOPTPASERKIALFTVNAKAVISAIDA	240
QY	241	DTIYRIPLLEREGLDLVDQLRLDVPADLSAMKQVVDGLHPTDRAVLAIVGKYVDH	300
DB	241	DTIYRIPLLEREGLDLVDQLRLDVPADLSAMKQVVDGLHPTDRAVLAIVGKYVDH	300
QY	301	TDVAKSLNEALIHGITHRKYVOISYIDSETTEARGTAKLVNDALIVPGSGRGVGEK	360
DB	301	TDVAKSLNEALIHGITHRKYVOISYIDSETTEARGTAKLVNDALIVPGSGRGVGEK	360
QY	361	ISTYRFARENKPIYLGICLQKQSAVIFPARNVGLGASHTEFLPKSPHYIGIITMD	420
DB	361	ISTYRFARENKPIYLGICLQKQSAVIFPARNVGLGASHTEFLPKSPHYIGIITMD	420

Db 361 ISTVQYARBNKIPYIGICLGMQVAVIYARNVIGSDANSTFDSKSGHPVALLTTEWQD 420

QY 421 EAGEIVTRDESDLDGGTMRGAOKCRILKADSLAFOLYQKOVITERRRHRYEENNOYLKOL 480

Db 421 AAGATEIRFEASDLGGTMRGAQEQOLQGTGLVHDCADVIVERRHRYEENNNMLLPOL 480

QY 481 EAAAGKFSKSLDGRVLEIIELEPHEWFLACQPHPEFTSPRNGHALPSGFTVAAAKH 538

Db 481 EAGAGKISGRSDGALVEVEAPEHWPVACQPHPEFTSPRNGHPSGFTVAAALMKY 538

## RESULT 2

AD0859

CTP synthetase (imported) - Salmonella enterica subsp. enterica serovar Typhi (strain CT C/Specties: Salmonella enterica subsp. enterica serovar Typhi

A/Note: this species has also been called Salmonella typhi

C/Date: 09-Nov-2001 #sequence\_revision 09-Nov-2001 #text\_change 18-Nov-2002

C/Accession: AD0859

R/Parthill, J.; Dougan, G.; James, K.D.; Thomson, N.R.; Pickard, D.; Main, J.; Churcher, th. T.; Comerton, P.; Cronin, A.; Davis, P.; Davies, R.M.; Dowd, L.; White, N.; Farrar, S.; Koul, S.; O'Garra, P.

Nature 413, 848-852, 2001

A/Authors: Parry, C.; Quail, M.; Rutherford, K.; Simmonds, M.; Skelton, J.; Stevens, K.; A/Title: Complete genome sequence of a multiple drug resistant Salmonella enterica serov

A/Reference number: AB0502; MUID:21534947; PMID:11677608

A/Accession: AD0859

A/Status: preliminary

A/Molecule type: DNA

A/Residues: 1-545 <PAR>

A/Cross-References: GB:AL513382; PIDN:CAD06059.1; PID:g16504026; GSPDB:GN00176

A/Genetics: STY3082

C/Superfamily: CTP synthase

Query Match 69.8%; Score 1946; DB 2; Length 545;

Best Local Similarity 68.2%; Pred. No. 3e-117;

Matches 367; Conservative 73; Mismatches 98; Indels 0; Gaps 0;

2 TKFIFITGVVSSLGKGIASSLAILEDGKLTITKLDPIYINVDGTPMSPOGHEV 61

3 TNYIFVITGVVSSLGKGIASSLAILEDGKLTITKLDPIYINVDGTPMSPOGHEV 62

62 TEDGATDLDGHERFLKTYTKKNNFTTGOVYEOVLENERKDYLGATVQVPHITDE 121

63 TEDGATDLDGHERFLKTYTKKNNFTTGOVYEOVLENERKDYLGATVQVPHITNA 122

122 IKRIVYSAEGKQVALLIEVGTVGDIESTLPLETRQMGVLEGRDALFIHILTVPIYKS 181

123 IKRIVYSAEGKQVALLIEVGTVGDIESTLPLETRQMGVLEGRDALFIHILTVPIYAA 182

182 AGEVTKPTQHSVKELRTIGIOPDILICRSEQIPASERKIALFTVAERKAVISAIDAD 241

183 AGEVTKPTQHSVKELRTIGIOPDILICRSEQIPASERKIALFTVAERKAVISMKDVD 242

242 TYIRIPLLREOGDLDVQDLRDVPAADLSAMKRVVDGLTHPTDEVSAIVGKYVDHT 301

243 STYKIPOLISQSGDLDVQDLRDVPAADLSAMKRVVDGLTHPTDEVSAIVGKYVDHT 302

302 DAYKSLNEALIHAGIHTTRKVOISYIDSETTEAGETAKLNVDAILVPGGGERGVSKI 361

303 DAYKSLNEALIHAGIHTTRKVOISYIDSETTEAGETAKLNVDAILVPGGGERGVSKI 362

362 STVPRARENKIPYIGICLGMQVAVIYARNVIGSDANSTFDSKSGHPVALLTTEWQD 421

363 STVPRARENKIPYIGICLGMQVAVIYARNVIGSDANSTFDSKSGHPVALLTTEWQD 422

422 AGEIVTRDESDLDGGTMRGAOKCRILKADSLAFOLYQKOVITERRRHRYEENNOYLKOL 481

423 AGEIVTRDESDLDGGTMRGAOKCRILKADSLAFOLYQKOVITERRRHRYEENNOYLKOL 482

482 AAGKFSKSLDGRVLEIIELEPHEWFLACQPHPEFTSPRNGHALPSGFTVAAAKH 539

483 AAGKFSKSLDGRVLEIIELEPHEWFLACQPHPEFTSPRNGHALPSGFTVAAAKH 540

## RESULT 3

E82074

CTP synthase VC2448 (imported) - Vibrio cholerae (strain N16961 serogroup O1)

C/Specties: Vibrio cholerae

C/Date: 18-Aug-2000 #sequence\_revision 20-Aug-2000 #text\_change 02-Feb-2001

C/Accession: E82074

R/Heidelberg, J.F.; Eisen, J.A.; Nelson, W.C.; Clayton, R.A.; Gwin, M.L.; Dodson, R.J.;

1. R.R.; Mekalanos, J.J.; Venter, J.C.; Fraser, C.M.

Nature 406 477-483, 2000

A/Title: DNA Sequence of both chromosomes of the cholera pathogen Vibrio cholerae.

A/Reference number: AB0305; MUID:20406833; PMID:10952301

A/Accession: E82074

A/Status: preliminary

A/Molecule type: DNA

A/Residues: 1-545 <HRI>

A/Cross-References: GB:AE004315; GB:AE003852; MUID:9657024; PIDN:AAE95590.1; GSPDB:GN001

A/Experimental source: serogroup O1; strain N16961; biotype El Tor

A/Genetics: VC2448

A/Map position: 1

C/Superfamily: CTP synthase

Query Match 69.0%; Score 1924; DB 2; Length 545;

Best Local Similarity 67.5%; Pred. No. 7.7e-116;

Matches 362; Conservative 76; Mismatches 98; Indels 0; Gaps 0;

2 TKFIFITGVVSSLGKGIASSLAILEDGKLTITKLDPIYINVDGTPMSPOGHEV 61

3 TNYIFVITGVVSSLGKGIASSLAILEDGKLTITKLDPIYINVDGTPMSPOGHEV 62

62 TEDGATDLDGHERFLKTYTKKNNFTTGOVYEOVLENERKDYLGATVQVPHITDE 121

63 TEDGATDLDGHERFLKTYTKKNNFTTGOVYEOVLENERKDYLGATVQVPHITNA 122

122 IKRIVYSAEGKQVALLIEVGTVGDIESTLPLETRQMGVLEGRDALFIHILTVPIYKS 181

123 IKRIVYSAEGKQVALLIEVGTVGDIESTLPLETRQMGVLEGRDALFIHILTVPIYAA 182

182 AGEVTKPTQHSVKELRTIGIOPDILICRSEQIPASERKIALFTVAERKAVISAIDAD 241

183 AGEVTKPTQHSVKELRTIGIOPDILICRSEQIPASERKIALFTVAERKAVISMKDVD 242

242 TYIRIPLLREOGDLDVQDLRDVPAADLSAMKRVVDGLTHPTDEVSAIVGKYVDHT 301

243 STYKIPOLISQSGDLDVQDLRDVPAADLSAMKRVVDGLTHPTDEVSAIVGKYVDHT 302

302 DAYKSLNEALIHAGIHTTRKVOISYIDSETTEAGETAKLNVDAILVPGGGERGVSKI 361

303 DAYKSLNEALIHAGIHTTRKVOISYIDSETTEAGETAKLNVDAILVPGGGERGVSKI 362

362 STVPRARENKIPYIGICLGMQVAVIYARNVIGSDANSTFDSKSGHPVALLTTEWQD 421

363 STVPRARENKIPYIGICLGMQVAVIYARNVIGSDANSTFDSKSGHPVALLTTEWQD 422

422 AGEIVTRDESDLDGGTMRGAOKCRILKADSLAFOLYQKOVITERRRHRYEENNOYLKOL 481

423 AGEIVTRDESDLDGGTMRGAOKCRILKADSLAFOLYQKOVITERRRHRYEENNOYLKOL 482

482 AAGKFSKSLDGRVLEIIELEPHEWFLACQPHPEFTSPRNGHALPSGFTVAAAKH 537

483 AAGKFSKSLDGRVLEIIELEPHEWFLACQPHPEFTSPRNGHALPSGFTVAAAKH 538

## RESULT 4

SYCTR

CTP synthase (BC 6.3.4.2) (validated) - Escherichia coli (strain K-12)

N/Alternate names: CTP-synthetase; UTP-ammonia ligase

C/Specties: Escherichia coli

C/Date: 31-Mar-1988 #sequence\_revision 21-Nov-1997 #text\_change 01-Mar-2002

C/Accession: H65059; A25608

R;Blattner, F.R.; Plunkett III, G.; Bloch, C.A.; Perna, N.T.; Burland, V.; Riley, M.; Rose, D.J.; Mau, B.; Shao, Y.  
 Science 277, 1453-1462, 1997  
 A>Title: The complete genome sequence of *Escherichia coli* K-12.  
 A;Reference number: A64720; MUID:9742617; PMID:9278503  
 A;Accession: H65059  
 A;Status: preliminary; nucleic acid sequence not shown; translation not shown  
 A;Molecule type: DNA  
 A;Residues: 1-545 <BLAT>  
 A;Cross-references: GB:AE000361; GB:U00096; NID:92367160; PIDN:AACT5822.1; PID:91789142  
 A;Experimental source: strain K-12, substrain MG1655  
 R;Meng, M.; Makarov, C.A.; Zalkin, H.  
 J. Biol. Chem. 261, 5568-5574, 1986  
 A>Title: Nucleotide sequence of *Escherichia coli* pyrg encoding CTP synthetase.  
 A;Reference number: A92584; MUID:86168304; PMID:3514618  
 A;Accession: A25608  
 A;Molecule type: DNA  
 A;Residues: 1337,'L',339-475,'S',477-488,'RA',491-545 <MEN>  
 A;Cross-references: GB:M12843  
 C;Genetics:  
 A;Gene: pyrg  
 A;Map position: 60 min  
 C;Function: homodimer; aggregates to a tetramer  
 A;Description: this glutamine amidotransferase catalyzes the terminal reaction in the de  
 C;Keywords: homodimer; 1-gase; pyrimidine nucleotide biosynthesis  
 F;2-545/Product: CTP synthase #status experimental <MAT>

Query Match 68.7%; Score 1916; DB 1; Length 545;  
 Best Local Similarity 67.7%; Pred. No. 2.5e-115;

Matches 363; Conservative 71; Mismatches 102; Indels 0; Gaps 0;

2 TKEFTTGGVSSLSKGIASLSAILEDGKVTITKLDPIYNDPGTMSFPGHGVFV 61  
 3 TTYIFVTGGVSSLSKGIASLSAILEDGKVTITKLDPIYNDPGTMSFPGHGVFV 62  
 62 TEDGATDLDGHTERFRTKMTKNNFTTGOVYEQVLENERKDYLGATVQVPHITDE 121  
 63 TEDGATDLDGHTERFRTKMTKNNFTTGOVYEQVLENERKDYLGATVQVPHITDE 122  
 122 IKRRVYASAEKQVALIEVGTGDIESTPLETTRQMGVLEGRDALPHILTVPYKLS 181  
 123 IKRRVYASAEKQVALIEVGTGDIESTPLETTRQMGVLEGRDALPHILTVPYKLS 182  
 182 AGELETKPTQSHSVKELRTIGIOPDILICRSEQPIPAERKRIALPTNVAEKAVISAIDAD 241  
 183 AGELETKPTQSHSVKELRTIGIOPDILICRSEQPIPAERKRIALPTNVAEKAVISAIDAD 242  
 242 TTYRTPILIRQGLDVLVDQRLDVPADLSAEKQVVDGLTHPDEVISAIVGKYVDHT 301  
 243 TTYRTPILIRQGLDVLVDQRLDVPADLSAEKQVVDGLTHPDEVISAIVGKYVDHT 302  
 302 DAYKSINAEALHAGITRHKQVSIYDSETEIEAGTAKLKNVDALIVPGGGERGVGKI 361  
 303 DAYKSINAEALHAGITRHKQVSIYDSETEIEAGTAKLKNVDALIVPGGGERGVGKI 362  
 362 STVPARENKIPLYGICLQMSAVIEFPAENVGLSEGSTFEFLPKSPPVGLITERNDE 421  
 363 STVPARENKIPLYGICLQMSAVIEFPAENVGLSEGSTFEFLPKSPPVGLITERNDE 422  
 422 AGELETKPTQSHSVKELRTIGIOPDILICRSEQPIPAERKRIALPTNVAEKAVISAIDAD 481  
 423 AGELETKPTQSHSVKELRTIGIOPDILICRSEQPIPAERKRIALPTNVAEKAVISAIDAD 482  
 482 AAGMKFSGKSLDGLVEILIEPEHWPFLACQFHEPFTSTPRNGHALPFGFVBAAK 537  
 483 AAGMKFSGKSLDGLVEILIEPEHWPFLACQFHEPFTSTPRNGHALPFGFVBAAK 538

RESULT 5  
 H91083  
 CTP synthetase [imported] - *Escherichia coli* (strain 0157:H7, substrain RIMD 0509952)

C;Species: *Escherichia coli*  
 C;Date: 18-Jul-2001 #sequence\_revision 18-Jul-2001 #text\_change 03-Aug-2001  
 C;Accession: H91083  
 R;Hayashi, T.; Makino, K.; Ohnishi, M.; Kurokawa, K.; Ishii, K.; Yokoyama, K.; Han, C.G.;  
 Gasekawa, N.; Yasunaga, T.; Kuhara, S.; Shiba, T.; Hattori, M.; Shinagawa, H.  
 DNA Res. 8, 11-22, 2001  
 A>Title: Complete genome sequence of enterohemorrhagic *Escherichia coli* O157:H7 and genom  
 A;Reference number: A99629; MUID:21156231; PMID:11258796  
 A;Accession: H91083  
 A;Status: preliminary  
 A;Molecule type: DNA  
 A;Residues: 1-545 <HAY>  
 A;Cross-references: GB:BA000007; PIDN:BA837063.1; PID:G13363111; GSPDB:GN00154  
 A;Experimental source: strain 0157:H7, substrain RIMD 0509952  
 C;Genetics:  
 A;Gene: EC93640  
 C;Superfamily: CTP synthase

Query Match 68.7%; Score 1916; DB 2; Length 545;  
 Best Local Similarity 67.7%; Pred. No. 2.5e-115;

Matches 363; Conservative 71; Mismatches 102; Indels 0; Gaps 0;

2 TKEFTTGGVSSLSKGIASLSAILEDGKVTITKLDPIYNDPGTMSFPGHGVFV 61  
 3 TTYIFVTGGVSSLSKGIASLSAILEDGKVTITKLDPIYNDPGTMSFPGHGVFV 62  
 62 TEDGATDLDGHTERFRTKMTKNNFTTGOVYEQVLENERKDYLGATVQVPHITDE 121  
 63 TEDGATDLDGHTERFRTKMTKNNFTTGOVYEQVLENERKDYLGATVQVPHITDE 122  
 122 IKRRVYASAEKQVALIEVGTGDIESTPLETTRQMGVLEGRDALPHILTVPYKLS 181  
 123 IKRRVYASAEKQVALIEVGTGDIESTPLETTRQMGVLEGRDALPHILTVPYKLS 182  
 182 AGELETKPTQSHSVKELRTIGIOPDILICRSEQPIPAERKRIALPTNVAEKAVISAIDAD 241  
 183 AGELETKPTQSHSVKELRTIGIOPDILICRSEQPIPAERKRIALPTNVAEKAVISAIDAD 242  
 242 TTYRTPILIRQGLDVLVDQRLDVPADLSAEKQVVDGLTHPDEVISAIVGKYVDHT 301  
 243 TTYRTPILIRQGLDVLVDQRLDVPADLSAEKQVVDGLTHPDEVISAIVGKYVDHT 302  
 302 DAYKSINAEALHAGITRHKQVSIYDSETEIEAGTAKLKNVDALIVPGGGERGVGKI 361  
 303 DAYKSINAEALHAGITRHKQVSIYDSETEIEAGTAKLKNVDALIVPGGGERGVGKI 362  
 362 STVPARENKIPLYGICLQMSAVIEFPAENVGLSEGSTFEFLPKSPPVGLITERNDE 421  
 363 STVPARENKIPLYGICLQMSAVIEFPAENVGLSEGSTFEFLPKSPPVGLITERNDE 422  
 422 AGELETKPTQSHSVKELRTIGIOPDILICRSEQPIPAERKRIALPTNVAEKAVISAIDAD 481  
 423 AGELETKPTQSHSVKELRTIGIOPDILICRSEQPIPAERKRIALPTNVAEKAVISAIDAD 482  
 482 AAGMKFSGKSLDGLVEILIEPEHWPFLACQFHEPFTSTPRNGHALPFGFVBAAK 537  
 483 AAGMKFSGKSLDGLVEILIEPEHWPFLACQFHEPFTSTPRNGHALPFGFVBAAK 538

# RESULT 6

CTP synthetase [imported] - *Escherichia coli* (strain 0157:H7, substrain EDL933)  
 C;Species: *Escherichia coli*  
 C;Date: 16-Feb-2001 #sequence\_revision 16-Feb-2001 #text\_change 14-Sep-2001  
 C;Accession: A85929  
 R;Perna, N.T.; Plunkett III, G.; Burland, V.; Mau, B.; Glasner, J.D.; Rose, D.J.; Mayhew,  
 Miller, L.; Grobeck, E.J.; Davis, A.; Dimalanta, E.; Potamotis, K.; Apodaca,  
 Nature 409, 529-533, 2001  
 A>Title: Genome sequence of enterohemorrhagic *Escherichia coli* O157:H7.  
 A;Reference number: A85480; MUID:21074935; PMID:11206551  
 A;Accession: A85929  
 A;Status: preliminary  
 A;Molecule type: DNA

A:Residues: 1-545 <STO>  
 A:Cross-references: GB:AE005174; NID:G15212252; PIDN:AA057893.1; GSPDB:GN00145; UWGP:240  
 A:Experimental source: strain 0157:H7, substrain EDL933  
 C:Genetics:  
 A:Gene: *pyrG*  
 C:superfamily: CTP synthase

Query Match 68.7%; Score 1916; DB 2; Length 545;  
 Best Local Similarity 67.7%; Pred. No. 2.5e-115;  
 Matches 363; Conservative 71; Mismatches 102; Indels 0; Gaps 0;

```

QY 2 TKEFIFITGVSSLSKGIASLSAIIEDRGKVTITKLDPIYINVDPGTMSFGHGEVYV 61
DB 3 TNYIFVTGVSSLSKGIASLSAIIEDRGKVTITKLDPIYINVDPGTMSFGHGEVYV 62
QY 62 TEDGAEITDLDGHERFLKRTMTKNNFTTGQVYEQVLSNERKGDYLGATVQVPHITDE 121
DB 63 TEDGAEITDLDGHERFLKRTMTKNNFTTGQVYEQVLSNERKGDYLGATVQVPHITDE 122
QY 122 IKRRVYESAGKDVALLIVGGTVGDIISLPLETTIRQMGVGLGRDALFIHLTLVPIYKS 181
DB 123 IKRRVYESAGKDVALLIVGGTVGDIISLPLETTIRQMGVGLGRDALFIHLTLVPIYKS 182
QY 182 AGELTKRTQMSVKEKRTIGQPDLICRSQPIPASRRKIALFTVAEKAVISALDAD 241
DB 183 SGVVKTKPTQSHVKEKRTIGQPDLICRSQPIPASRRKIALFTVAEKAVISALDAD 242
QY 242 TTYRIPLLRRCGLDVLVDQLRLDVPADLSAMEKVVDGLTPTDEVSIAIVKTYVDHT 301
DB 243 SIYKIRGLKSGLDVYICRPSLCPBEANLAEWQVLYEESNPGEGYTIQMGKVELP 302
QY 302 DAYKSLNFBALIHAGIHTRHVKQIYSTIDSETTEABGTAKLVNDAILVPGGFGYRGVEK 361
DB 303 DAYKSLNFBALIHAGIHTRHVKQIYSTIDSETTEABGTAKLVNDAILVPGGFGYRGVEK 362
QY 362 STVRFARENKIPLYGICLGMQSAVIEFARNVGLGASHTFELPKSPHPVIGLITEMDE 421
DB 363 TTRAFARENKIPLYGICLGMQSAVIEFARNVGLGASHTFELPKSPHPVIGLITEMDE 422
QY 422 AGELVTRDESDSLGTMELGAOKRLKADSLAFOLYQKDVITERRHRRYEFNNQYKOLE 481
DB 423 NGNVEVRSKSDJGTMELGAOKRLKADSLAFOLYQKDVITERRHRRYEFNNQYKOLE 482
QY 482 AAGMKFSKSLDGLRVEIIEHPHPWFLACQHPHPTSTPRNGHALPSGFEVAAAK 537
DB 483 DAGIRAVAGSGDDQLVRIIVPMHPWVACQHPHPTSTPRNGHALPSGFEVAAAK 538

```

## RESULT 7

AC0410  
 CTP synthase (EC 6.3.4.2) [imported] - Yersinia pestis (strain CO92)  
 C:Species: Yersinia pestis  
 C>Date: 02-Nov-2001 #sequence\_revision 02-Nov-2001 #text\_change 27-Nov-2001  
 C:Accession: AC0410  
 R:Perkhill, J.; Wren, B.W.; Thomson, N.R.; Titchell, R.W.; Holden, M.T.G.; Prentice, M.B.;  
 deno-1ariga, A.M.; Chillingworth, T.; Cronin, A.; Davies, R.M.; Davis, P.; Dougan, G.;  
 11, M.; Rutherford, K.; Simmonds, M.; Skelton, J.; Stevens, K.; Whitehead, S.; Barrett,  
 Nature 413, 523-527, 2001  
 A:Title: Genome sequence of Yersinia pestis, the causative agent of plague.  
 A:Reference number: AB0001; MUID:21470413; PMID:11586360  
 A:Accession: AC0410  
 A:Status: Preliminary  
 A:Molecule type: DNA  
 A:Residues: 1-545 <KUR>  
 A:Cross-references: GB:AL590842; PIDN:CAC92607.1; PID:G15981303; GSPDB:GN00175  
 C:Genetics:  
 A:Gene: *pyrG*  
 C:superfamily: CTP synthase  
 C:Keywords: ligase

Query Match 68.5%; Score 1909; DB 2; Length 545;  
 Best Local Similarity 66.2%; Pred. No. 7.1e-115;  
 Matches 355; Conservative 80; Mismatches 102; Indels 0; Gaps 0;

```

QY 2 TKEFIFITGVSSLSKGIASLSAIIEDRGKVTITKLDPIYINVDPGTMSFGHGEVYV 61
DB 3 TNYIFVTGVSSLSKGIASLSAIIEDRGKVTITKLDPIYINVDPGTMSFGHGEVYV 62
QY 62 TEDGAEITDLDGHERFLKRTMTKNNFTTGQVYEQVLSNERKGDYLGATVQVPHITDE 121
DB 63 TEDGAEITDLDGHERFLKRTMTKNNFTTGQVYEQVLSNERKGDYLGATVQVPHITDE 122
QY 122 IKRRVYESAGKDVALLIVGGTVGDIISLPLETTIRQMGVGLGRDALFIHLTLVPIYKS 181
DB 123 IKRRVYESAGKDVALLIVGGTVGDIISLPLETTIRQMGVGLGRDALFIHLTLVPIYKS 182
QY 182 AGELTKRTQMSVKEKRTIGQPDLICRSQPIPASRRKIALFTVAEKAVISALDAD 241
DB 183 SGVVKTKPTQSHVKEKRTIGQPDLICRSQPIPASRRKIALFTVAEKAVISALDAD 242
QY 242 TTYRIPLLRRCGLDVLVDQLRLDVPADLSAMEKVVDGLTPTDEVSIAIVKTYVDHT 301
DB 243 SIYKIRGLKSGLDVYICRPSLCPBEANLAEWQVLYEESNPGEGYTIQMGKVELP 302
QY 302 DAYKSLNFBALIHAGIHTRHVKQIYSTIDSETTEABGTAKLVNDAILVPGGFGYRGVEK 361
DB 303 DAYKSLNFBALIHAGIHTRHVKQIYSTIDSETTEABGTAKLVNDAILVPGGFGYRGVEK 362
QY 362 STVRFARENKIPLYGICLGMQSAVIEFARNVGLGASHTFELPKSPHPVIGLITEMDE 421
DB 363 TTRAFARENKIPLYGICLGMQSAVIEFARNVGLGASHTFELPKSPHPVIGLITEMDE 422
QY 422 AGELVTRDESDSLGTMELGAOKRLKADSLAFOLYQKDVITERRHRRYEFNNQYKOLE 481
DB 423 NGNVEVRSKSDJGTMELGAOKRLKADSLAFOLYQKDVITERRHRRYEFNNQYKOLE 482
QY 482 AAGMKFSKSLDGLRVEIIEHPHPWFLACQHPHPTSTPRNGHALPSGFEVAAAKH 539
DB 483 DAGIRAVAGSGDDQLVRIIVPMHPWVACQHPHPTSTPRNGHALPSGFEVAAAGDYQ 540

```

## RESULT 8

F64181  
 CTP synthase (EC 6.3.4.2) - Haemophilus influenzae (strain Rd KW20)  
 C:Species: Haemophilus influenzae  
 C>Date: 18-Aug-1995 #sequence\_revision 18-Aug-1995 #text\_change 16-Jul-1999  
 C:Accession: F64181  
 R:Feilschmann, R.D.; Adams, M.D.; White, O.; Clayton, R.A.; Kirkness, E.F.; Kerlavage, A.;  
 R:Feilschmann, R.D.; Adams, M.D.; White, O.; Clayton, R.A.; Kirkness, E.F.; Kerlavage, A.;  
 Gocayne, J.D.; Scott, J.; Shirley, J.; Liu, L.; Glodok, A.; Kelley, J.W.; Weidman, J.;  
 D.M.; Brandon, R.C.; Fine, L.D.; Fritschman, J.L.; Fuhrmann, J.L.; Geoghagen, N.S.M.  
 Science 269, 495-512, 1995  
 A:Authors: Gnehm, C.L.; McDonald, L.A.; Small, K.V.; Fraser, C.M.; Smith, H.O.; Venter,  
 A:Title: Whole-genome random sequencing and assembly of Haemophilus influenzae Rd.  
 A:Reference number: A64000; MUID:95350630; PMID:7542800  
 A:Accession: F64181  
 A:Status: nucleic acid sequence not shown; translation not shown  
 A:Molecule type: DNA  
 A:Residues: 1-545 <TIG>  
 A:Cross-references: GB:U32788; GB:I42023; NID:G1574629; PIDN:AA022733.1; PID:G1574630; 1  
 C:superfamily: CTP synthase  
 C:Keywords: ligase; pyrimidine nucleotide biosynthesis

Query Match 68.5%; Score 1908.5; DB 2; Length 545;  
 Best Local Similarity 67.5%; Pred. No. 7.6e-115;  
 Matches 364; Conservative 71; Mismatches 103; Indels 1; Gaps 1;

```

QY 2 TKEFIFITGVSSLSKGIASLSAIIEDRGKVTITKLDPIYINVDPGTMSFGHGEVYV 61
DB 3 TNYIFVTGVSSLSKGIASLSAIIEDRGKVTITKLDPIYINVDPGTMSFGHGEVYV 62
QY 62 TEDGAEITDLDGHERFLKRTMTKNNFTTGQVYEQVLSNERKGDYLGATVQVPHITDE 121
DB 63 TEDGAEITDLDGHERFLKRTMTKNNFTTGQVYEQVLSNERKGDYLGATVQVPHITDE 122
QY 122 IKRRVYESAGKDVALLIVGGTVGDIISLPLETTIRQMGVGLGRDALFIHLTLVPIYKS 181
DB 123 IKRRVYESAGKDVALLIVGGTVGDIISLPLETTIRQMGVGLGRDALFIHLTLVPIYKS 182
QY 182 AGELTKRTQMSVKEKRTIGQPDLICRSQPIPASRRKIALFTVAEKAVISALDAD 241
DB 183 SGVVKTKPTQSHVKEKRTIGQPDLICRSQPIPASRRKIALFTVAEKAVISALDAD 242
QY 242 TTYRIPLLRRCGLDVLVDQLRLDVPADLSAMEKVVDGLTPTDEVSIAIVKTYVDHT 301
DB 243 SIYKIRGLKSGLDVYICRPSLCPBEANLAEWQVLYEESNPGEGYTIQMGKVELP 302
QY 302 DAYKSLNFBALIHAGIHTRHVKQIYSTIDSETTEABGTAKLVNDAILVPGGFGYRGVEK 361
DB 303 DAYKSLNFBALIHAGIHTRHVKQIYSTIDSETTEABGTAKLVNDAILVPGGFGYRGVEK 362
QY 362 STVRFARENKIPLYGICLGMQSAVIEFARNVGLGASHTFELPKSPHPVIGLITEMDE 421
DB 363 TTRAFARENKIPLYGICLGMQSAVIEFARNVGLGASHTFELPKSPHPVIGLITEMDE 422
QY 422 AGELVTRDESDSLGTMELGAOKRLKADSLAFOLYQKDVITERRHRRYEFNNQYKOLE 481
DB 423 NGNVEVRSKSDJGTMELGAOKRLKADSLAFOLYQKDVITERRHRRYEFNNQYKOLE 482
QY 482 AAGMKFSKSLDGLRVEIIEHPHPWFLACQHPHPTSTPRNGHALPSGFEVAAAKH 539
DB 483 DAGIRAVAGSGDDQLVRIIVPMHPWVACQHPHPTSTPRNGHALPSGFEVAAAGDYQ 540

```

Db 123 IKDRVIAAGQGHVIVVGVGTVDIESLPLEALRQLAUVNGREHTLFMELTLVPIPT 182  
 QY 182 AGELTKPTQHSYKELRTTIGIOPDILICRSEQIPASERRKIALFTNVAEKAVISALDA 241  
 Db 183 ABEVTKPTQHSYKELRTTIGIOPDILICRSDMIPTNERAKIALFTNVAERAVISIKDYN 242  
 QY 242 TTYRIPLLRBOGLDLDVVDQRLDVPADLISAMEKVDGLTHPDEVSIALVGVKVDHT 301  
 Db 243 STYQIPALLKSOGLDLPVCEFRITCPEDLTEMEDVLYKQANPVEGVITIGVGVKTELP 302  
 QY 302 DAVKSLNALIHAIGHTRKVOQSYIDSETEIABEPAKIKANDALIVPGFSEGVGKI 361  
 Db 303 DAKSVNELKHAAGLTKNELSVNKTIDSDVETKGVYVGLKIDGLIVPGFSEGVGKI 362  
 QY 362 STYFARENKIPYLGICLQKQSAVIEFARNVGLGASHTETLPKSPHPVIGLITEMDE 421  
 Db 363 RTQVARENKIPYLGICLQKQALIEYANVAGLTKANSSEFDKCEQVVALITEMODA 422  
 QY 422 AGEVLRDSDSDGGMTRIGAKCRKADSLAFOLYQKQVITERHRHRYEPNNYKOLE 481  
 Db 423 EKVTEVTRDSBDGGMTRIGASQCHLVSGSRARLKGKETTERHRHRYEPNNYKOLE 482  
 QY 482 AAGMKFSKSLDGRVLEIIELEPEHFWFLACQHPPEFTSPRNGHALFSGVEAA-AKHK 539  
 Db 483 KGLKATVGLSADKQVLEIIEVPHFWFVACQHPPEFTSPRNGHALFAGFVKAAYENHK 541

## RESULT 9

C82700  
 CTP synthetase XF1268 (imported) - Xylella fastidiosa (strain 9a5c)  
 C:Species: Xylella fastidiosa  
 C:Date: 18-Aug-2000 #sequence\_revision 20-Aug-2000 #text\_change 02-Sep-2000  
 C:Accession: C82700  
 R:Anonymous, The Xylella fastidiosa Consortium of the Organization for Nucleotide Sequen-  
 Nature 406, 151-157, 2000  
 A:Title: The genome sequence of the plant pathogen Xylella fastidiosa.  
 A:Reference number: A82515; PMID:20365717; PMID:10910347  
 A:Note: for a complete list of authors see reference number A59328 below  
 A:Accession: C82700  
 A:Status: preliminary  
 A:Molecule type: DNA  
 A:Residues: 1-544 <SIM>  
 A:Cross-references: GB:AE003962; GB:AE003849; NID:G9106270; PIDN:AAF84097.1; GSDDB:GN001  
 A:Experimental source: strain 9a5c  
 R:Simpson, A.J.G.; Reinach, F.C.; Arruda, P.; Abreu, F.A.; Acencio, M.; Alvaranga, R.; A-  
 Bionnes, M.R.S.; Bueno, M.R.P.; Camargo, A.A.; Canario, L.E.A.; Carraro, D.M.; Carver, H-  
 as-Neco, E.; Docena, C.; El-Dorri, H.; Facincani, A.P.; Ferreira, A.J.S.  
 Submitted to Genbank, June 2000  
 A:Authors: Ferreira, V.C.A.; Ferro, J.A.; Freaga, J.S.; Franca, S.C.; Franco, M.C.; Frohm-  
 J.D.; Junqueira, M.L.; Kemper, E.L.; Klitařma, J.P.; Klieger, J.E.; Kuramae, E.E.; Laig-  
 Chado, M.A.; Madeira, A.M.B.N.; Madeira, H.M.F.; Marino, C.L.; Marques, M.V.; Martins, E-  
 A:Authors: Martins, E.M.F.; Matsubara, A.Y.; Menck, C.F.M.; Miracca, B.C.; Miyaki, D.Y-  
 F.G.; Nunes, L.R.; Oliveira, M.A.; de Oliveira, M.C.; de Oliveira, R.C.; Palmeri, D.Y-  
 Rodrigues, V.; Rosa, A.J. de M.; de Rosa Jr., V.E.; de Sa, R.G.; Santelli, R.V.; Sawaak  
 A:Authors: da Silva, A.C.R.; da Silva, F.R.; da Silva, M.M.; Silva Jr., M.A.; da Silva  
 M.; Tshako, M.H.; Vallada, H.; Van Sluys, M.A.; Verjovski-Almeida, S.; Vettore, A.L.; Z  
 A:Reference number: A59328  
 A:Contents: annotation  
 C:Genetics:  
 A:Gene: XF1268  
 C:Superfamily: CTP synthase

Query Match 68.3%; Score 1905; DB 2; Length 554;  
 Best Local Similarity 66.1%; Pred. No. 1,3e-114;  
 Matches 361; Conservative 79; Mismatches 102; Indels 4; Gaps 2;  
 QY 1 MKKFLITGVSVSSIGKGIASASLAILEDRLKVTITKLDPIYNDGTMSPFOHGEVF 60  
 Db 1 MKKFLITGVSVSSIGKGIASASLAILEDRLKVTITKLDPIYNDGTMSPFOHGEVF 60  
 QY 61 VTDEGAETDLDLGHYERFLKTTMTKNNFTTGQVYEQVLANERKGDYLGATVQVPHITD 120  
 Db 61 VTDEGAETDLDLGHYERFLKTTMTKNNFTTGQVYEQVLANERKGDYLGATVQVPHITD 120  
 QY 61 VTDEGAETDLDLGHYERFLKTTMTKNNFTTGQVYEQVLANERKGDYLGATVQVPHITD 120  
 Db 61 VTDEGAETDLDLGHYERFLKTTMTKNNFTTGQVYEQVLANERKGDYLGATVQVPHITD 120

QY 121 EIKRNVESAAGKQVALTEGVGTVDIESLPLETRQMGVELGRDALFHTLTVPIK 180  
 Db 121 EIRCIDATATASPDVAIVIGTVDIESLPLEIRQVRIERGAEKRTMFMTLTVPIA 180  
 QY 181 SAGELTKPTQHSYKELRTTIGIOPDILICRSEQIPASERRKIALFTNVAEKAVISALDA 240  
 Db 181 ABEVTKPTQHSYKELRTTIGIOPDILICRSDMIPTNERAKIALFTNVAERAVISIKDYN 240  
 QY 241 PTYRIPLLRBOGLDLDVVDQRLDVPADLISAMEKVDGLTHPDEVSIALVGVKVDHT 298  
 Db 241 DVLVGMPELRRQGLDEIVDQFLKSTGASLADISEMEDVDVADKHPDEVTIAYGVK 300  
 QY 299 DHTDAKSLNALIHAIGHTRKVOQSYIDSETEIABEPAKIKANDALIVPGFSEGVGKI 358  
 Db 301 DYQDYKVGALKKGGRGKQRTKVLKWEKQDLGSDMGALKIDGLIVPGFSEGVGKI 360  
 QY 359 GKISTVARENKIPYLGICLQKQSAVIEFARNVGLGASHTETLPKSPHPVIGLITEM 418  
 Db 361 GKVLASRYARQRPVYRGICYQKQAAVDYARHVAAGLEGASTENDRQSPHPVIALITEM 420  
 QY 419 MDEAGELVRDSDSDGGMTRIGAKCRKADSLAFOLYQKQVITERHRHRYEPNNYK 478  
 Db 421 RTTGSEVERRDEKSDIGCTMELIGQEQRLKAGTLVRELVGSDVVGERRHRYEPNNYK 480  
 QY 479 QLEAAGMKFSKSLDGRVLEIIELEPEHFWFLACQHPPEFTSPRNGHALFSGVEAA 536  
 Db 481 QLEDAAGLVIAKSMDDTLVEMILPREHFWFLACQHPPEFTSPRNGHALFAGFVKAAR 540  
 QY 537 KHKTOG 542  
 Db 541 ARKAGG 546

## RESULT 10

F81798  
 CTP synthase (EC 6.3.4.2) NMA1742 (imported) - Neisseria meningitidis (strain Z2491 serog-  
 C:Species: Neisseria meningitidis  
 C:Date: 05-May-2000 #sequence\_revision 05-May-2000 #text\_change 02-Feb-2001  
 C:Accession: F81798  
 R:Parkhill, J.; Achtman, M.; James, K.D.; Bentley, S.D.; Churcher, C.; Klee, S.R.; Morell  
 ; Holtroyd, S.; Jagsis, K.; Leather, S.; Moule, S.; Mungall, K.; Quail, M.A.; Rajandream,  
 Nature 404, 502-506, 2000  
 A:Title: Complete DNA sequence of a serogroup A strain of Neisseria meningitidis Z2491.  
 A:Reference number: A81775; PMID:20222556; PMID:10761919  
 A:Accession: F81798  
 A:Status: preliminary  
 A:Molecule type: DNA  
 A:Residues: 1-544 <PAR>  
 A:Cross-references: GB:AL162757; GB:AL157959; NID:G7380371; PIDN:CA884970.1; PID:G7380384  
 A:Experimental source: serogroup A, strain Z2491  
 C:Genetics:  
 A:Gene: pyrg; NMA1742  
 C:Superfamily: CTP synthase  
 C:Keywords: ligase

Query Match 67.9%; Score 1892.5; DB 2; Length 544;  
 Best Local Similarity 66.1%; Pred. No. 8.1e-114;  
 Matches 358; Conservative 80; Mismatches 101; Indels 3; Gaps 2;  
 QY 1 MKKFLITGVSVSSIGKGIASASLAILEDRLKVTITKLDPIYNDGTMSPFOHGEVF 60  
 Db 1 MKKFLITGVSVSSIGKGIASASLAILEDRLKVTITKLDPIYNDGTMSPFOHGEVF 60  
 QY 61 VTDEGAETDLDLGHYERFLKTTMTKNNFTTGQVYEQVLANERKGDYLGATVQVPHITD 120  
 Db 61 VTDEGAETDLDLGHYERFLKTTMTKNNFTTGQVYEQVLANERKGDYLGATVQVPHITD 120  
 QY 121 EIKRNVESAAGKQVALTEGVGTVDIESLPLETRQMGVELGRDALFHTLTVPIK 180  
 Db 121 EIKRNVESAAGKQVALTEGVGTVDIESLPLETRQMGVELGRDALFHTLTVPIK 180  
 QY 181 SAGELTKPTQHSYKELRTTIGIOPDILICRSEQIPASERRKIALFTNVAEKAVISALDA 240  
 Db 181 SAGELTKPTQHSYKELRTTIGIOPDILICRSDMIPTNERAKIALFTNVAERAVISIKDYN 240





N:Alternate names: CTP-synthetase; UTP-ammocinia ligase  
 C:Species: *Bacillus subtilis*  
 C:Date: 30-Jun-1992 #sequence revision 30-Jun-1992 #text change 16-Jun-2000  
 C:Accession: A32354; S55423; G69610  
 R:Trach, K.; Chapman, J.M.; Pigott, P.; Lecocq, D.; Hoch, J.A.  
 J. Bacteriol. 170, 4194-4208, 1988  
 A:Title: Complete sequence and transcriptional analysis of the spoOF region of the *Bacillus subtilis* chromosome  
 A:Accession: A32354  
 A:Reference number: A31883; PMID:88314920; PMID:2457578  
 A:Molecule type: DNA  
 A:Residues: 1-535 <TRA>  
 A:Cross-references: GB:M22039; NID:G460310; PIDN:AA16801.1; PID:G143597  
 R:Glaeser, P.; Danchin, A.  
 Submitted to the EMBL Data Library, May 1995  
 A:Description: Cloning and sequencing of the *Bacillus subtilis* chromosomal region from 3  
 A:Reference number: S55414  
 A:Molecule type: DNA  
 A:Accession: S55423  
 A:Residues: 1-535 <GLA>  
 A:Cross-references: EMBL:249782; NID:G853752; PIDN:CA89870.1; PID:G853762  
 R:Kumst, F.; Ogatawa, N.; Moszer, I.; Albertini, A.M.; Alloni, G.; Azevedo, V.; Berter  
 C.; Bron, S.; Brouillet, S.; Brusch, C.V.; Caldwell, B.; Capuano, V.; Carter, N.M.; Cho  
 A.; Ehrlich, S.D.; Emerson, P.T.; Ehtan, K.D.; Errington, J.; Fabret, C.; Ferrari, B.  
 Nature 390, 249-256, 1997  
 A:Authors: Foulger, D.; Filtz, C.; Fujita, M.; Fujita, Y.; Fuma, S.; Galizzi, A.; Gallier  
 iech, J.; Harwood, C.R.; Henaut, A.; Hilbert, H.; Holsappel, S.; Hosono, S.; Hullo, M.F.  
 Koester, P.; Korungstein, G.; Krogh, S.; Kumano, M.; Kurita, K.; Lapidus, A.; Lardinois,  
 A:Authors: Lauber, J.; Lazarevic, V.; Lee, S.M.; Levine, A.; Liu, H.; Masuda, S.; Mauei  
 Y, M.; Ogawa, K.; Ogatawa, A.; Odega, B.; Park, S.H.; Parro, V.; Pohl, T.M.; Portetel  
 Rieger, M.; Rivolta, C.; Rocha, E.; Roche, B.; Rose, M.; Sadate, Y.; Sato, T.; Scanlon,  
 A:Authors: Schleif, S.; Schreier, R.; Scelfone, F.; Sekiguchi, J.; Sekowska, A.; Serot  
 akouch, M.; Tanakoshi, A.; Tanaka, T.; Tefstera, P.; Tognoni, A.; Tosato, V.; Uchiyama,  
 T.; Winters, P.; Wipac, A.; Yamamoto, H.; Yamane, K.; Yasunoto, K.; Yata, K.; Yoshida, K  
 A:Authors: Yoshikawa, H.F.; Zumbstein, E.; Yoshikawa, H.; Danchin, A.  
 A:Title: The complete genome sequence of the Gram-positive bacterium *Bacillus subtilis*.  
 A:Reference number: A65580; PMID:98044033; PMID:9384377  
 A:Accession: C69610  
 A:Status: nucleic acid sequence not shown; translation not shown  
 A:Molecule type: DNA  
 A:Residues: 1-535 <RNA>  
 A:Cross-references: GB:299123; GB:A1009126; NID:G2636240; PIDN:CA15743.1; PID:G2636252  
 A:Experimental source: strain 168  
 C:Comment: This enzyme is a glutamine amidotransferase that catalyzes the terminal react  
 C:Genetics:  
 A:Gene: ctpA; pyrG  
 A:Map position: 37 min  
 C:Superfamily: CTP synthase  
 C:Keywords: ligase; pyrimidine nucleotide biosynthesis

Query Match 58.2%; Score 1622.5; DB 1; Length 535;  
 Best Local Similarity 57.0%; Pred. No. 1.7e-96;  
 Matches 310; Conservative 94; Mismatches 127; Indels 13; Gaps 3;

QY 1 MTKFIIITGGVSSSLCKGIAASLAILLEDRGLKVTITLDPYINVDPTGMSPPQHGVP 60  
 DB 1 MTKYIFVTGGVSSSLCKGIVAAASLGLLNKRLNLTITQKDPYINVDPTGMSPPQHGVP 60  
 QY 61 VTEDGAEITDLDGHEVRFKTKMTKKNFTTQVYEVLNERKGDYLGATVQVPHITD 120  
 DB 61 VTDDGAEITDLDGHEVRFIDINLKNFSNVTTKISTVLKKERRDYLGGTQVPHITN 120  
 QY 121 EIKRKYVSEAB--GKDVALIEVGGTYGDIESTLPLETTIRQMGVZLGRDRALEFHLITVY 178  
 DB 121 ELKDRYVYRAGKSTNADVITTEIGTVDIESLPLEAIRQMSDIDGRNVMYHCTLVY 180  
 QY 179 IKSAGELKTPQHSVKEITIGIOEDILICSEOPIPASEERKIALFTNVAEKAVISAI 238  
 DB 181 IKAAGELKTPQHSVKEITIGIOEDILICSEOPIPASEERKIALFTNVAEKAVISAI 240  
 QY 239 DADVTIRILLRBEGLDVLVDQLRDVPAADLSAMEKVDGLTHPTDEVSIAIVSKY 298  
 DB 241 DADNLYSIFLELOKGLDVLVDQLRDVPAADLSAMEKVDGLTHPTDEVSIAIVSKY 300

QY 299 DETDAYNSLNEALIHAGIHTRHKVOISYIDSETIEAGTAXL-KNVDAIVPGGPGRGV 357  
 DB 301 ELPPAYISVSESLRHAGYAFEDTVKVMKINAEETENNIAELTSGTGIIIVPGGPGRGV 360  
 QY 358 ECKSTYRFAENKIPYIGICLGWASVTEFARVVLGQHSFELPKSPHPIGLITE 417  
 DB 361 ECKVATYKARENKIPYIGICLGWASVTEFARVVLGQHSFELPKSPHPIGLITE 420  
 QY 418 MWDSAGELVTRDESDIDGTRLGAKCRKLKAUSLAEQLYQKDYITERHRRHYEFNNQYL 477  
 DB 421 QKDY-----EDIGTFLRLGLYPCXKLEGTAKAEVYQDEVVYRHHRRHYEFNNIEFR 470  
 QY 478 KQLAAGKRSRSGSLDGRVHITLPEHPPTLACQHPPTSTPRNHALFSGFVEZAAK 537  
 DB 471 QQMEEGFVFSGTSPDRLVETIILKDPFWVASQFHPFSRPTRPDPLKFGTASVE 530  
 QY 538 HKTQ 541  
 DB 531 AANQ 534

Search completed: January 29, 2004, 15:55:33  
 Job time : 17.0063 secs

GenCore version 5.1.6  
Copyright (c) 1993 - 2004 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: January 29, 2004, 15:41:08 ; Search time 10.244 Seconds

(without alignments) 2497.314 Million cell updates/sec

Title: US-09-941-947a-16

Sequence: 1 MTKFIFITGVVSLGKGLA.....HALFGFVFAAKKTKGTGA 544

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 127863 seqs, 47026705 residues

Total number of hits satisfying chosen parameters: 127863

Minimum DB seq length: 0

Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : SwissProt\_41.\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

## SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	2008	75.0	542	1	PYRG_PSEAE
2	1948	69.9	544	1	PYRG_NITRU
3	1946	68.8	544	1	PYRG_SALTU
4	1938	69.5	544	1	PYRG_XANCP
5	1937	69.3	546	1	PYRG_VIBPA
6	1931	69.3	554	1	PYRG_XANAC
7	1925	69.0	545	1	PYRG_VIBVU
8	1924	69.0	544	1	PYRG_VIBCH
9	1918	68.8	542	1	PYRG_PASXU
10	1916	68.7	544	1	PYRG_ECOLI
11	1909	68.5	544	1	PYRG_YERPE
12	1908.5	68.5	545	1	PYRG_HAEIN
13	1905	68.3	554	1	PYRG_XYLPA
14	1892.5	67.9	544	1	PYRG_NEIMA
15	1888.5	67.7	544	1	PYRG_NEIMA
16	1822	65.4	554	1	PYRG_PALSO
17	1737	62.3	545	1	PYRG_BUCAI
18	1729.5	62.0	553	1	PYRG_BUCAP
19	1685.5	60.5	544	1	PYRG_BUCBP
20	1662	59.6	552	1	PYRG_WIGER
21	1645.5	58.8	532	1	PYRG_BACBD
22	1640.5	58.8	537	1	PYRG_THERN
23	1639	58.8	531	1	PYRG_AOUNA
24	1622.5	58.2	535	1	PYRG_BACSU
25	1615.5	57.9	535	1	PYRG_CLOAB
26	1603.5	57.5	552	1	PYRG_SYNT3
27	1600.5	57.4	532	1	PYRG_LISIN
28	1600.5	57.4	530	1	PYRG_LISMO
29	1595	57.2	550	1	PYRG_CAUCR
30	1592	57.1	542	1	PYRG_BRUNE
31	1586	56.9	542	1	PYRG_CLOPE
32	1583.5	56.8	535	1	PYRG_CLOPE
33	1580.5	56.7	546	1	PYRG_SYN7

34	1576	56.5	536	1	PYRG_STAKX
35	1574.5	56.5	545	1	PYRG_ANASP
36	1563.5	56.1	540	1	PYRG_METUA
37	1562	56.0	535	1	PYRG_STRPN
38	1562	56.0	542	1	PYRG_AGRFS
39	1555.5	55.8	537	1	PYRG_PYRAB
40	1555	55.8	542	1	PYRG_RHILU
41	1552.5	55.7	535	1	PYRG_METKA
42	1542	55.3	544	1	PYRG_AZOBK
43	1540.5	55.3	537	1	PYRG_PYRPU
44	1538.5	55.2	537	1	PYRG_PYRPO
45	1507	54.1	543	1	PYRG_CAMJB

## ALIGNMENTS

RESULT 1  
ID PYRG\_PSEAE STANDARD, PRT, 542 AA.  
AC Q9HXZ4;  
DT 28-FEB-2003 (Rel. 41, Created)  
DT 28-FEB-2003 (Rel. 41, Last sequence update)  
DT 28-FEB-2003 (Rel. 41, Last annotation update)  
DE CTP synthase (EC 6.3.4.2) (UTP--ammonia ligase) (CTP synthetase).  
GN PYRG OR PA3637.  
OS Pseudomonas aeruginosa.  
OC Bacteria; Proteobacteria; Gammaproteobacteria; Pseudomonadales;  
OC Pseudomonadaceae; Pseudomonas.  
OX NCBI\_TaxID=287;  
RN [1]  
RP SEQUENCE FROM N.A.  
RC STRAIN=ATCC 15692 / PA01;  
RX MEDLINE=20437337; PubMed=10964043;  
RA Stover C.K., Pham X.-Q.T., Erwin A.L., Mizoguchi S.D., Warren P.,  
RA Hickey M.J., Brinkman F.S.L., Hutnagle W.O., Kowalik D.J., Lagrou M.,  
RA Gardner R.L., Goltz L., Tolentino E., Westbrook-Wadman S., Yagou M.,  
RA Brody L.L., Coulter S.N., Folger K.R., Kas A., Larbig K., Lam R.M.,  
RA Smith K.A., Spencer D.H., Wong G.K.-S., Mu Z., Paulsen I.T.,  
RA Reizer J., Sailer M.H., Hancock R.E.W., Lory S., Olson M.V.;  
RA "Complete genome sequence of Pseudomonas aeruginosa PA01, an  
RT opportunistic pathogen.";  
RT Nature 406:959-964(2000).  
RL -1- FUNCTION: Catalyzes the ATP-dependent amination of UTP to CTP with  
either L-glutamine or ammonia as the source of nitrogen (By  
similarity).  
CC -1- CATALYTIC ACTIVITY: ATP + UTP + NH(3) = ADP + phosphate + CTP.  
CC -1- ENZYME REGULATION: Allosterically activated by GTP, when glutamine  
is the substrate. Inhibited by CTP (By similarity).  
CC -1- PATHWAY: Pyrimidine biosynthesis; conversion of UMP to CTP, third  
(last) step.  
CC -1- SUBUNIT: Homotrimer (By similarity).  
CC -1- SIMILARITY: Belongs to the CTP synthase family.  
CC -1- SIMILARITY: Contains 1 type-1 glutamine amidotransferase domain.  
CC This Swiss-Prot entry is copyright. It is produced through a collaboration  
between the Swiss Institute of Bioinformatics and the EMBL outstation -  
the European Bioinformatics Institute. There are no restrictions on its  
use by non-profit institutions as long as its content is in no way  
modified and this statement is not removed. Usage by and for commercial  
entities requires a license agreement (See <http://www.ebi.ac.uk/announcements>  
or send an email to [license@ebi.ac.uk](mailto:license@ebi.ac.uk)).  
CC -----  
CC EMBL, AB004783; AB007025.1; -;  
CC PIR, B83192; B83192.  
CC HAMAP, MF\_01227; -1.  
CC InterPro, IPR000991; GATase\_1.  
CC InterPro, IPR004468; PyrG.  
CC Pfam, PF00117; GATase\_1.  
CC TIGRFAMs, TIGR00337; PyrG\_1.  
CC PROSITE, PS00442; GATASE\_TYPE\_1; 1.  
CC Pyrimidine biosynthesis; Ligase; Glutamine amidotransferase;  
CC Complete proteome.



Db 361 IMASVARNHRIFPYIGICIGCMQAVIIEFRNRNLCJENNAHSTIEFDQDPYPLZGILITMRD  
Qy 421 EAGELVTRDESDUDGCTMFLGAKCKLADSLAFLQYQDVITVERHRRRYFFNNQYIKOL  
Db 421 RCRCVRKRSKQDITDLGGCTMRLLGGQECILKPKHTLAHRIYADQKVIETHRRRRYVNAEPIPOL  
Qy 481 EAAQKMFSGKSLDGRVLEIIEIP--EHMFPLTACQHPETSPPRGNALFGFGEAA 535  
Db 481 EQAQKMHISGLSAGEDLCMIETLPEQSEHPKFTVACQHPERTISTPRNGHPLFASYIOAA 480  
481 EQAQKMHISGLSAGEDLCMIETLPEQSEHPKFTVACQHPERTISTPRNGHPLFASYIOAA 537

RESULT 3  
PYRG\_SALTY  
ID PYRG\_SALTY STANDARD; PRT; 544 AA.  
AC OXENE; 28-FEB-2003 (Rel. 41, Created)  
DT 28-FEB-2003 (Rel. 41, Last sequence update)  
DT 15-SEP-2003 (Rel. 42, Last annotation update)  
DE CYP synthase (EC 6.3.4.2) (UTP--ammonia ligase) (CTP synthetase).  
GN PYRG OR STM2953 OR STY3082 OR T2854.  
OS Salmonella typhimurium, and  
OC Salmonella typh.  
OC Bacteria; Proteobacteria; Gammaproteobacteria; Enterobacteriales;  
OC Enterobacteriaceae; Salmonella.  
OX NCBI\_TaxID=602, 601;  
RN [1]  
RP SEQUENCE FROM N.A.  
RC SPECIES=S.typhimurium; STRAIN=L72 / SGSC1412 / ATCC 700720;  
RX MEDLINE=21534948; PubMed=11677609;  
RA McClelland M., Sanderson K.E., Spatch J., Clifton S.W., Latreille P.,  
RA Courtney L., Porwollik S., Ali O., Dante M., Du P., Hou S., Layman D.,  
RA Leonard S., Nguyen C., Scott K., Holmes A., Grewal N., Mulvaney E.,  
RA Ryan E., Sun H., Flores L., Miller W., Stoneking T., Nhan M.,  
RA Waterston R., Wilson R.K.;  
RT "complete genome sequence of Salmonella enterica serovar Typhimurium  
LT2.";  
RL Nature 413:852-856(2001).  
RN [2]  
RP SEQUENCE FROM N.A.  
RC SPECIES=S.typh; STRAIN=CT18;  
RX MEDLINE=21534947; PubMed=11677608;  
RA Partholl J., Dougan G., James K.D., Thomson N.R., Pickard D., Wain J.,  
RA Churcher C., Mangall K.L., Bentley S.D., Holden M.T.G., Sebaiha M.,  
RA Baker S., Basham D., Brooks K., Chillingworth T., Connor P.,  
RA Cronin A., Davis P., Davies R.M., Dowd L., White N., Farrar J.,  
RA Felwell T., Hamlin N., Haque A., Hien T.T., Holroyd S., Jajels K.,  
RA Krogh A., Larsen T.S., Leather S., Moule S., O'Garra P., Parry C.,  
RA Quail M., Rutherford K., Simmonds M., Skelton J., Stevens K.,  
RA Whitehead S., Barrall B.G.;  
RT "complete genome sequence of a multiple drug resistant Salmonella  
RT enterica serovar Typh CT18.";  
RL Nature 413:848-852(2001).  
RN [3]  
RP SEQUENCE FROM N.A.  
RC SPECIES=S.typh; STRAIN=Ty2 / ATCC 700931;  
RX MEDLINE=22531367; PubMed=1264504;  
RA Burgard W., Liou S.-R., Plunkett G. III, Mayhew G.F., Rose D.J.,  
RA Deng L.V., Kodoyianni V., Schwartz D.C., Blatner F.R.;  
RT "Comparative genomics of Salmonella enterica serovar Typhi strains Ty2  
and CT18.";  
RL J. Bacteriol. 185:2330-2337(2003).  
CC -1- FUNCTION: Catalyzes the ATP-dependent amination of UMP to CTP with  
CC either L-glutamine or ammonia as the source of nitrogen (by  
CC similarity).  
CC -1- CATALYTIC ACTIVITY: ATP + UTP + NH(3) = ADP + phosphate + CTP.  
CC -1- ENZYME REGULATION: Allosterically activated by GTP, when glutamine  
CC is the substrate. Inhibited by CTP (by similarity).  
CC -1- PATHWAY: Pyrimidine biosynthesis; conversion of UMP to CTP; third  
CC (last) step.  
CC -1- SUBUNIT: Hemotetramer (by similarity).  
CC -1- SIMILARITY: Belongs to the CTP synthase family.  
CC -1- SIMILARITY: Contains 1 type-I glutamine amidotransferase domain.

```

CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
CC
DR EMBL; AE008835; CAL21833.1; -
DR EMBL; AL627276; CAD06059.1; -
DR EMBL; AE016843; AO070411.1; -
DR StyGene; SG32272; Pyrg.
DR HAMAP; MF_01227; -; 1.
DR InterPro; IPR000931; GATase_1.
DR Pfam; PF00117; GATase; 1.
DR PROSITE; PS00442; GATASE_TYPE_I; 1.
DR Pyrimidine biosynthesis; Ligase; Glutamine amidotransferase;
KM Complete proteome.
KV
FT INIT MET 0
FT DOMAIN 1 299 BY SIMILARITY.
FT FT 300 544 AMINATOR DOMAIN.
FT FT 300 544 GLUTAMINE AMIDOTRANSFERASE.
FT ACT_SITE 378 378 GATASE (BY SIMILARITY).
FT ACT_SITE 514 514 GATASE (BY SIMILARITY).
FT ACT_SITE 516 516 GATASE (BY SIMILARITY).
FT ACT_SITE 516 516 GATASE (BY SIMILARITY).
CC SEQUENCE 544 AA; 59990 MM; E4BCA885F342EAB8 CRC64;

```

Query Match	69.8%	Score 1946;	DB 11;	Length 544;
Best Local Similarity	68.2%	Pred. No. 1	4e-114;	
Matches 367;	Conservative 98;	Mismatches 98;	Indels 0;	Gaps 0;

QY	2	TKFIFITGVVSSIGKIDIAASLAAIEDGKLTITKIDLPYINVDGTVSPPHGEFV	61
Db	2	TNIYFVIGGVSSIGKIDIAASLAAIEANGLVITIMKIDPIINVDGTVSPHIGEFV	61
QY	62	TEDGAETDLDLGHYERFLKTTTKANNFTTQYIY3OVLNBERKCDYIGATYQVIFPHITDE	121
Db	62	TEDGAETDLDLGHYERFIRTEMSSRNFTTGRISYDYLRRBRDDYIGATYQVIFPHITNA	121
QY	122	IKRIVYSAGKDVALIEVGATGVDIESLPLEFTRQMGVELGDRALFHLFTLVPIYKS	181
Db	122	IKERVYEGGBHDVYVIEIGTQVDIESLPLEFLIRDLANDIGREHLLFHLFTLVPIYLA	181
QY	182	AGELKTPYHSHVKELTFTIGIOPDILICRSEOPIPASERRKIALFTVAKRAYSALDAD	241
Db	182	AGEKTPYHSHVKELTISIGIOPDILICRSDRVPANERAKIALPCVNPBEKAVISMDDVD	241
QY	242	TIYVILPILLRBQGLDLDVNDQLRDVPAALISAMEKVVDDGLTHPTDEVSIAIVGKYDHT	301
Db	242	STYKIPILLSSQGLDVTCKRPSLNCERANLSERKQVITYEANRAGVITIMGKCYIELP	301
QY	302	DAVKSINLEALIHAGIMRHRKVOISYIDSETIIEAGTAKLNVDAILVPGGFBGRGVBEKI	361
Db	302	DAVKSIVTEALHGKGLKRRVTYNIKILIDSOVETRGVEILDLDAILLPGGFGYGVBEKI	361
QY	362	STVRAPEENKIPYIGICIGMOSAVIEPARVVGIEGASHTEPLKPSHPVITGLITTEWMD	421
Db	362	ATARYAERENKIPYIGICISQVALIEPARVVGADGNANSTIEVPDCKTPVVALITTEWRD	421
QY	422	AGELVTRDEDSDILOGITNELCAQCRLEKADSLAFOLYOKDYITERHRRRYEPFNQYLKQLE	481
Db	422	DGANEVSRBSKSDLGITNRLCAQCCOLSDSDSLVQLYQASTYIVERHRRRYEVNNMLLKQIB	481
QY	482	AAGKKEFGSKLDGLGVYIIELPEHPWFLAOCFHEPFTSPRNGALTSFGFVEAAKAKR	539
Db	482	AAGLRVAGRGDDQVLBIIEVPHNPFVACQFHEPFTSTRODGHPLRAGFVAKANENQ	539

DT 28-FEB-2003 (Rel. 41, Last sequence update)  
 DT 28-FEB-2003 (Rel. 41, Last annotation update)  
 DB CTP synthase (EC 6.3.4.2) (UTP--ammonia ligase) (CTP synthetase).  
 GN PYRG OR XCC1697.  
 OS Xanthomonas campestris (pv. campestris).  
 OC Bacteria; Proteobacteria; Gammaproteobacteria; Xanthomonadales;  
 CC Xanthomonadaceae; Xanthomonas.  
 NX NCBI\_TaxID=340;  
 RN  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=ATCC 33913 / NCPB 528;  
 RX MEDLINE=22022145; PubMed=12024217;  
 RA de Silva A.C.R., Ferro J.A., Reinach P.C., Farah C.S., Burlan L.R.,  
 RA Quaggio R.B., Monteiro-Vitorello C.B., Van Sluys M.A., Almeida N.F.,  
 RA Alves L.M.C., do Amaral A.M., Bertolini M.C., Camargo L.B.A.,  
 RA Camarotte G., Camarvan F., Cardoso J., Chamberg F., Clapina L.P.,  
 RA Cicarelli R.M.B., Coutinho L.L., Cursino-Santos J.R., El-Doroty H.,  
 RA Faria J.B., Ferreira A.J.S., Ferreira R.C.C., Ferro M.I.T.,  
 RA Fontighieri E.F., Franco M.C., Greggio C.C., Guber A.,  
 RA Katsuyama A.M., Kishi L.T., Leite R.P., Lemos E.G.M., Lemos M.V.F.,  
 RA Locati E.C., Machado M.A., Madeira A.M.B.N., Martinez-Rossi N.M.,  
 RA Martins E.C., Meladas J., Menck C.F.M., Miyaki C.Y., Moon D.H.,  
 RA Moreira L.M., Novo M.T.M., Okura V.K., Oliveira M.C., Oliveira V.R.,  
 RA Pereira H.A., Rossi A., Sana J.A.D., Silva C., de Souza R.F.,  
 RA Spicola L.A.F., Takita M.A., Tamura R.E., Teixeira R.C., Tezza R.I.D.,  
 RA Trindade dos Santos M., Truffi D., Tsai S.M., White F.F.,  
 RA Serraldi J.C., Kitejima J.P.,  
 RA "Comparison of the genomes of two Xanthomonas pathogens with differing  
 RT host specificities."  
 RT Nature 417:459-463(2002).  
 RL -1- FUNCTION: Catalyzes the ATP-dependent amination of UTP to CTP with  
 CC either L-glutamine or ammonia as the source of nitrogen (3y  
 CC similarity).  
 CC -1- CATALYTIC ACTIVITY: ATP + UTP + NH(3) = ADP + phosphate + CTP.  
 CC -1- ENZYME REGULATION: Allosterically activated by GTP, when glutamine  
 CC is the substrate. Inhibited by CTP (By similarity).  
 CC -1- PATHWAY: Pyrimidine biosynthesis; conversion of UMP to CTP; third  
 CC (last) step.  
 CC -1- SUBUNIT: Homotetramer (By similarity).  
 CC -1- SIMILARITY: Belongs to the CTP synthase family.  
 CC -1- SIMILARITY: Contains 1 type-1 glutamine amidotransferase domain.  
 CC  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
 CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
 CC  
 CC EMBL; AE012271; AA040991.1; ALT\_INIT.  
 DR HAMAP; MF\_01227; .1.  
 DR InterPro; IPR000091; GATase\_1.  
 DR InterPro; IPR004468; Pyrg.  
 DR Pfam; PF00117; GATase; 1.  
 DR TIGRfam; TIGR00337; Pyrg; 1.  
 DR PROSITE; PS00442; GATASB\_TFB\_I; 1.  
 DR Pyrimidine biosynthesis; Ligase; Glutamine amidotransferase;  
 KW Complete proteome.  
 FT DOMAIN 1 301 AMINATOR DOMAIN.  
 FT DOMAIN 302 554 GLUTAMINE AMIDOTRANSFERASE.  
 FT ACT\_SITE 380 380 GATASE (BY SIMILARITY).  
 FT ACT\_SITE 518 518 GATASE (BY SIMILARITY).  
 FT ACT\_SITE 520 520 GATASE (BY SIMILARITY).  
 SQ SEQUENCE 554 AA; A2D70CA2DBDF1AE0 CR664;  
 Query Match 69.5%; Score 1938; DB 1; Length 554;  
 Best Local Similarity 67.2%; Pred. No. 4,4e-114;  
 Matches 367; Conservative 80; Mismatches 95; Indels 4; Gaps 2;

QY 61 VTDEGAEITDIDLGHTYEFKLTMTKKNFTTGGVYEVOLNENKGYLCAVYVTHITD 120  
 DB 61 VTDDGAEITDIDLGHTYEFKLTMTKKNFTTGGVYEVOLNENKGYLCAVYVTHITD 120  
 QY 121 EIKRRVYSAEGKVALIEVGTVGDIIESLPFETTRQMGVEKGRDALFHTLTVYIK 180  
 DB 121 EIKRCIDEATGDFVALIEIGTVGDIIESLPFETTRQMGVEKGRDALFHTLTVYIK 180  
 QY 181 SAGELKTKPTQSHVKEIRTYIGIOPDILICRSEQPIPAEERKIALFTNVAEAVISAIDA 240  
 DB 181 SAGELKTKPTQSHVKEIRTYIGIOPDILICRSEQPIPAEERKIALFTNVAEAVISAIDA 240  
 QY 241 DTYIRIPIALLREGQDGLVYDQRL--DYPAAQLSMEKVVQDGLTPHDEVSIAIVGKY 298  
 DB 241 DTYIRIPIALLREGQDGLVYDQRL--DYPAAQLSMEKVVQDGLTPHDEVSIAIVGKY 298  
 QY 299 DHTDAVKSILNHALIHAGIHTRHAKVQISYIDSETIEAGTKLRNVDAILVPGGGERGVE 358  
 DB 301 DHQDAVSVAPLALHGHGLRQRTKNTLTWBAQDLESDMALQIDGILVPGGGERGFE 360  
 QY 359 GKISTVPAEAKNTPIYIGICMGQSAITEPARNVNGEGHSTPELPKSHPVIGLTTEW 418  
 DB 361 GKVTSKYABHHKVPYIGICMGQAAVADYARHVDLDAANSTENDROSPHPVIGLTTEW 420  
 QY 419 MDZAGELVTEDESDGLGTWELAGQCRKLKADSLAPQLYOKDVITERRHRYEFNNQYLK 478  
 DB 421 RTATGVEKVDKSDSLGTWELAGQCRKLKADSLAPQLYOKDVITERRHRYEFNNQYLK 480  
 QY 479 QLEPAKMGKSGKSIDGLVETIELPE--HMFPLAQCHPEPTSPRNGALPSEFVEAAA 536  
 DB 481 QLEBAGVIGCKSKSDTLVEMVELPRDTHPMFTLACQAHPELTPRGHPFLIGVAAAR 540  
 QY 537 KHKTOG 542  
 DB 541 EKKAGG 546  
 RESULT 5  
 PRG\_VIBPA STANDARD; PRT; 546 AA.  
 ID PRG\_VIBPA  
 AC 0871B9;  
 DT 15-SEP-2003 (Rel. 42, Created)  
 DT 15-SEP-2003 (Rel. 42, Last sequence update)  
 DT 15-SEP-2003 (Rel. 42, Last annotation update)  
 DB CTP synthase (EC 6.3.4.2) (UTP--ammonia ligase) (CTP synthetase).  
 GN PYRG OR VP2562.  
 OS Vibrio parahaemolyticus.  
 OC Bacteria; Proteobacteria; Gammaproteobacteria; Vibrionales;  
 CC Vibrionaceae; Vibrio.  
 CC NCBI\_TaxID=670;  
 RN  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=RIMD 221063 / Serotype O3:K6;  
 RX MEDLINE=22508454; PubMed=12620739;  
 RA Makino K., Oshima K., Kurokawa K., Yokoyama K., Uda T., Tagomori K.,  
 RA Iijima Y., Naito M., Nakano M., Yamashita A., Kubota Y., Kimura S.,  
 RA Yasunaga T., Honda T., Shinagawa H., Hattori K., Iida T.,  
 RT "Genome sequence of Vibrio parahaemolyticus: a pathogenic mechanism  
 RT distinct from that of V. cholerae."  
 RL Lancet 361:743-749(2003).  
 CC -1- FUNCTION: Catalyzes the ATP-dependent amination of UTP to CTP with  
 CC either L-glutamine or ammonia as the source of nitrogen (By  
 CC similarity).  
 CC -1- CATALYTIC ACTIVITY: ATP + UTP + NH(3) = ADP + phosphate + CTP.  
 CC -1- ENZYME REGULATION: Allosterically activated by GTP, when glutamine  
 CC is the substrate. Inhibited by CTP (By similarity).  
 CC -1- PATHWAY: Pyrimidine biosynthesis; conversion of UMP to CTP; third  
 CC (last) step.  
 CC -1- SUBUNIT: Homotetramer (By similarity).  
 CC -1- SIMILARITY: Belongs to the CTP synthase family.  
 CC -1- SIMILARITY: Contains 1 type-1 glutamine amidotransferase domain.

CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
 CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).

DR EMBL: A005082; BAC60825.1; -  
 DR HAMAP: MF01227; -  
 DR PROSITE: PS00442; GATASE\_TYPE\_1;  
 KW Pyrimidine biosynthesis; Ligase; Glutamine amidotransferase;  
 KM Complete proteome.  
 FT DOMAIN 1 300 AMINATOR DOMAIN  
 FT ACT SITE 301 546 GLUTAMINE AMIDOTRANSFERASE  
 FT ACT SITE 379 379 GATASE (BY SIMILARITY)  
 FT ACT SITE 515 515 GATASE (BY SIMILARITY)  
 FT ACT SITE 517 517 GATASE (BY SIMILARITY)  
 SQ SEQUENCE 546 AA; 60126 MW; E33C935EASB89E4 CRC64;

Query Match 69.5%; Score 1937; DB 1; Length 546;  
 Best Local Similarity 68.2%; Pred. No. 5e-116; Mismatches 96; Indels 0; Gaps 0;  
 Matches 366; Conservative 75;

2 TKRPIITGVVSSLGKGIASSIAALIEDRGKVTITKDPYINVDPGTMSPFQHGVEYV 61  
 3 TWYIFVTGVVSSLGKGIASSIAALIEDRGKVTITKDPYINVDPGTMSPFQHGVEYV 62  
 62 TEDGATDLDLGHYERFLTKTKKNTFTQGYEVOYLNERNKGYLGAIVYPIHTD 121  
 63 TEDGATDLDLGHYERFLTKTKKNTFTQGYEVOYLNERNKGYLGAIVYPIHTD 122  
 122 IKRRVYSAEGKQVALIEVGTVGDIESTLPLETRQMGVEGRDALPIHILTVPYKS 181  
 123 IKRRVYSAEGKQVALIEVGTVGDIESTLPLETRQMGVEGRDALPIHILTVPYKS 182  
 182 AGELTKPTQHSVKELRTGIDPILICSEOPIPASERKIALFTNAKQVISAIDAD 241  
 183 AGELTKPTQHSVKELRTGIDPILICSEOPIPASERKIALFTNAKQVISAIDAD 242  
 242 TYIRIPILIREGDDLVYDQRLVDPADLSAMEVYVUGLHPDEVSIAIVKGYVDHT 301  
 243 TYIRIPILIREGDDLVYDQRLVDPADLSAMEVYVUGLHPDEVSIAIVKGYVDHT 302  
 302 DAYKSLNEALIHAGIHTREKQVQISYIDSETIEABGTAKLKNVDAILVPGFGERGEGKI 361  
 303 DAYKSLNEALIHAGIHTREKQVQISYIDSETIEABGTAKLKNVDAILVPGFGERGEGKI 362  
 362 STVRFRERKIKITPLIGICLQMSAVIEFANVVGLEAGASTEFLEKSPREYVIGLITERWDE 421  
 363 STVRFRERKIKITPLIGICLQMSAVIEFANVVGLEAGASTEFLEKSPREYVIGLITERWDE 422  
 422 AGELTVTRDSDSLGGMRLGAOKCRKLKADSLAFOLYQKQVITERRHREYFNNQYIKOLE 481  
 423 TGVNVEERTSSDLGGMRGSLQCHLEKTKARELYGASTTERRHREYFNNQYIKOLE 482  
 482 AAGMTPSGGSLDGRUYETLEPEHPWFLACQTHPEFTSPRNGALLFGFTEAANH 538  
 483 KALGVSGSLADKLVEMLENPAPHPFVAQPHPTSTPRDGHPLFAGFVFAAQY 539

RESULT 6  
 PYRG\_XANAC STANDARD; PRT; 554 AA.

AC Q8PLS3;  
 DT 28-FEB-2003 (rel. 41, Created)  
 DT 28-FEB-2003 (rel. 41, Last sequence update)  
 DT 28-FEB-2003 (rel. 41, Last annotation update)  
 DE CTP synthase (EC 6.3.4.2) (UTP--ammonia ligase) (CTP synthetase).  
 GN PYRG OR XAC1716.  
 OS Xanthomonas axonopodis (pv. citri).  
 CC Bacteria; Proteobacteria; Gammaproteobacteria; Xanthomonadales;  
 CC Xanthomonadaceae; Xanthomonas.

OK NCBI\_TaxID=92829;

RP SEQUENCE FROM N.A.  
 RC STRAIN=306 / ATCC 13902 / XV 101;  
 RX MEDLINE=22022145; PubMed=12024217;  
 RA da Silva A.C.R., Ferro J.A., Reinach F.C., Parah C.S., Porlan L.R.,  
 RA Quaggio R.B., Monteiro-Vitorello C.B., Van Sluys M.A., Almeida N.F.,  
 RA Alves L.M.C., do Amaral A.M., Bertolini M.C., Camargo L.E.A.,  
 RA Camarotte G., Camarvan F., Cardoso J., Chamberg F., Clapina L.P.,  
 RA Ciccarelli R.M.B., Coutinho L.B., Cursino-Santos J.R., El-Dorri H.,  
 RA Faria J.B., Ferreira A.J.S., Ferreira R.C.C., Ferro M.I.T.,  
 RA Fournigier R.F., Franco M.C., Greggio C.C., Gruber A.,  
 RA Katsuyama A.M., Kishi L.T., Leite R.P., Lemos E.G.M., Lemos M.V.F.,  
 RA Locali E.C., Machado M.A., Madeira A.M.B.N., Martinez-Rossi N.M.,  
 RA Martins E.C., Meidanis U., Wenck C.F.W., Miyaki C.Y., Moon D.H.,  
 RA Moreira L.M., Novo M.T.M., Okura V.K., Oliveira M.C., Oliveira V.R.,  
 RA Pereira H.A., Rossi A., Sana J.A.D., Silva C., de Souza R.F.,  
 RA Spindola L.A.F., Takita M.A., Tamura R.B., Teixeira E.C., Texeira R.I.D.,  
 RA Trindade dos Santos M., Truffi D., Teal S.M., White F.F.,  
 RA Serebail J.C., Kikajima J.P.;  
 RA "Comparison of the genomes of two Xanthomonas pathogens with differing  
 RT host specificities."  
 RL Nature 417:459-463(2002).  
 CC -1- FUNCTION: Catalyzes the ATP-dependent amination of UTP to CTP with  
 CC either L-glutamine or ammonia as the source of nitrogen (By  
 CC similarity).  
 CC -1- CATALYTIC ACTIVITY: ATP + UTP + NH(3) = ADP + phosphate + CTP.  
 CC -1- ENZYME REGULATION: Allosterically activated by GTP, when glutamine  
 CC is the substrate. Inhibited by CTP (By similarity).  
 CC -1- PATHWAY: Pyrimidine biosynthesis; conversion of UMP to CTP, third  
 CC (last) step.  
 CC -1- SUBUNIT: Homotrimer (By similarity).  
 CC -1- SIMILARITY: Belongs to the CTP synthase family.  
 CC -1- SIMILARITY: Contains 1 type-1 glutamine amidotransferase domain.

CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
 CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).

DR EMBL: A0011804; AAM6583.1; -  
 DR HAMAP: MF\_01227; -  
 DR InterPro: IPR000991; GATase\_1.  
 DR InterPro: IPR004468; Pyrg.  
 DR Pfam: PF00117; GATase, 1.  
 DR TIGRFAMs: TIGR00337; Pyrg, 1.  
 DR PROSITE: PS00442; GATASE\_TYPE\_1;  
 DR Pyrimidine biosynthesis; Ligase; Glutamine amidotransferase;  
 KW Complete proteome.  
 FT DOMAIN 1 301 AMINATOR DOMAIN  
 FT ACT SITE 302 554 GLUTAMINE AMIDOTRANSFERASE  
 FT ACT SITE 380 380 GATASE (BY SIMILARITY)  
 FT ACT SITE 518 518 GATASE (BY SIMILARITY)  
 FT ACT SITE 520 520 GATASE (BY SIMILARITY)  
 SQ SEQUENCE 554 AA; 61482 MW; 80DD42A9DBA23596 CRC64;

Query Match 69.3%; Score 1931; DB 1; Length 554;  
 Best Local Similarity 67.0%; Pred. No. 1.2e-113; Mismatches 97; Indels 4; Gaps 2;  
 Matches 366; Conservative 79;

QY 1 MTKRPIITGVVSSLGKGIASSIAALIEDRGKVTITKDPYINVDPGTMSPFQHGVEY 60  
 DB 1 MTKRPIITGVVSSLGKGIASSIAALIEDRGKVTITKDPYINVDPGTMSPFQHGVEY 60  
 QY 61 VTEGATFDLDLGHYERFLTKTKKNTFTQGYEVOYLNERNKGYLGAIVYPIHTD 120  
 DB 61 VTEGATFDLDLGHYERFLTKTKKNTFTQGYEVOYLNERNKGYLGAIVYPIHTD 120  
 QY 121 EIKRKYSAEGKQVALIEVGTVGDIESTLPLETRQMGVEGRDALPIHILTVPYIK 180

```

DB 121 EIRRIDATAGFVALIEIGTVDISLPELNAIROVTEREKAEMPHLITVPIA 180
QY 181 SAGELKTPGSHVXELTIGIOPDILICRSEOPIPASERRKIALFTVAEKAVISALDA 240
DB 181 AAGELKTPGSHVXELTIGIOPDILICRSEOPIPASERRKIALFTVNSRAVISCIDI 240
QY 241 DTVIPLILLRQGLDILVVOPLR--DYPADUSAMERKVDGLTHPDEVSIAIVKRY 298
DB 241 DVLVGMPELRLQGLDELVIDPQKLDKVAADISEMAVAVDAVGHPLDEVNIAVGVKY 300
QY 239 DHTDAKSLNEMALIHAGIHTHRKVOISYIDSETTEAEETAKLKNVDAILVGGFGEKVE 358
DB 301 DRQDAKVSAAELBRGHRQRKTKVNLKMLBAQDLGSGMSALQIDILVGGEDRFE 360
QY 359 GKISTVRARENKILYLGICLQMSAVIERANVVGLEGAISTEPKSPHPVIGLITW 418
DB 361 GKVOQSKVAREQKEYFGICYGMOAAVVDYAHVADLDNAASTENDRQSPHPVIGLITW 420
QY 419 MDAGELVTRDESDLGGMRLGAQCRKADSLAFOLYOKDVTERRHREYEFNNQYLK 478
DB 421 RTAGEVERKEDKSLDGLGTMRLGLOEORLKRGLARVYGVADVAERHREYEFNNQYLK 480
QY 479 QLEAAGMKFSGKSLDGLVETIELPE--HPWELACQFPEFTSTPRNGHALFSGFVEAA 536
DB 481 QLEDAGLVIGSKSMDDTVVEVVELPRDTHPWLACQAHPEELSTPRDGHPLFIFGVRAAR 540
QY 537 KHKQOG 542
DB 541 EKKAGG 546

```

RESULT 7  
ID PYRG VIBVU STANDARD; PRT; 545 AA.

```

AC 08DC63;
DT 15-SEP-2003 (Rel. 42, Created)
DT 15-SEP-2003 (Rel. 42, Last sequence update)
DT 15-SEP-2003 (Rel. 42, Last annotation update)
DB CTP synthase (EC 6.3.4.2) (UTP--ammonia ligase) (CTP synthetase).
GN PYRG OR VV11578.
OS Vibrio vulnificus.
OC Bacteria; Proteobacteria; Gammaproteobacteria; Vibrionales;
Vibrionaceae; Vibrrio.
OX NCBI_TaxID=672;
RN RP SEQUENCE FROM N.A.
RA STRAIN=CMCP6;
RA Rhee J.H., Kim S.Y., Chung S.S., Kim J.J., Moon Y.H., Jeong H.,
Choy H.E.;
RT "Complete genome sequence of Vibrio vulnificus CMCP6."
RL Submitted (DEC-2002) to the EMBL/GenBank/DBJ databases.
CC -1- FUNCTION: Catalyzes the ATP-dependent amination of UTP to CTP with
either L-glutamine or ammonia as the source of nitrogen (By
similarity).
CC -1- CATALYTIC ACTIVITY: ATP + UTP + NH(3) = ADP + phosphate + CTP.
CC -1- ENZYME REGULATION: Allosterically activated by GTP, when glutamine
is the substrate. Inhibited by CTP (By similarity).
CC -1- PATHWAY: Pyrimidine biosynthesis; conversion of UMP to CTP; third
(last) step.
CC -1- SUBUNIT: Homodimer (By similarity).
CC -1- SIMILARITY: Belongs to the CTP synthase family.
CC -1- SIMILARITY: Contains 1 type-1 glutamine amidotransferase domain.
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
between the Swiss Institute of Bioinformatics and the EMBL outstation-
the European Bioinformatics Institute. There are no restrictions on its
use by non-profit institutions as long as its content is in no way
modified and this statement is not removed. Usage by and for commercial
entities requires a license agreement (See http://www.ebi.ac.uk/announcements/
or send an email to license@ebi.ac.uk).
CC EMBL; AF016802; AAC10002.1; -
CC HAMAP; MF_01227; -; 1.

```

```

DR InterPro; IPR000991; GATase_1.
DR InterPro; IPR004468; Pyrg.
DR Pfam; PF00117; GATase; 1.
DR TIGRPFAMs; TIGR00337; Pyrg; 1.
DR PROSITE; PS00442; GATase type I; 1.
KW Pyrimidine biosynthesis; Ligase; Glutamine amidotransferase;
Complete proteome.
FT DOMAIN 1 300
FT ACT_SITE 301 545
FT ACT_SITE 379 379
FT ACT_SITE 515 515
FT ACT_SITE 517 517
SQ SEQUENCE 545 AA; 59872 MW; 0E71A40595F2805 CRC64;

```

Query Match 69.0%; Score 1925; DB 1; Length 545;  
Best local similarity 67.0%; Pred. No. 2, 8e-113;  
Matches 360; Conservative 80; Mismatches 97; Indels 0; Gaps 0;

```

QY 2 TKPFTIGGVVSSLSKGLAASLAILEDRLKVTITKLDPIYINVPQTMSPFQHGVEV 61
DB 3 TNYIFVGVVSSLSKGLAASLAILEDRLKVTITKLDPIYINVPQTMSPFQHGVEV 62
QY 62 TEDGATDLDLGHYERFLTKTTRKNNFTQGVYBOYLNERKGDYIATVQVIPHITDE 121
DB 63 TEDGATDLDLGHYERFLTKTTRKNNFTQGVYBOYLNERKGDYIATVQVIPHITND 122
QY 122 IKRVYSSAGKQVALIEVGGTVDIESLPETLTIDQVGLGRDRALFHLTLVPYKS 181
DB 123 IKRVYSSAGKQVALIEVGGTVDIESLPETLTIDQVGLGRDRALFHLTLVPYLA 182
QY 182 AGEIKTPGSHVXELTIGIOPDILICRSEOPIPASERRKIALFTVAEKAVISALDA 241
DB 183 AGEIKTPGSHVXELTIGIOPDILICRSEOPIPASERRKIALFTVAEKAVISALDA 242
QY 242 TIRIPILLRQGLDILVVOPLR--DYPADUSAMERKVDGLTHPDEVSIAIVKRYDHT 301
DB 243 TIRIPILLRQGLDILVVOPLR--DYPADUSAMERKVDGLTHPDEVSIAIVKRYDHT 302
QY 302 DAVKSLNEMALIHAGIHTHRKVOISYIDSETTEAEETAKLKNVDAILVGGFGEKVEK 361
DB 303 DAVKSLNEMALIHAGIHTHRKVOISYIDSETTEAEETAKLKNVDAILVGGFGEKVEK 362
QY 362 STVRARENKILYLGICLQMSAVIERANVVGLEGAISTEPKSPHPVIGLITWMD 421
DB 363 LAATYARENKILYLGICLQMSAVIERANVVGLEGAISTEPKSPHPVIGLITWMD 422
QY 422 AGEIKTPGSHVXELTIGIOPDILICRSEOPIPASERRKIALFTVAEKAVISALDA 481
DB 423 AGEIKTPGSHVXELTIGIOPDILICRSEOPIPASERRKIALFTVAEKAVISALDA 482
QY 482 AAGKTSKSLDGLVETIELPE--HPWELACQFPEFTSTPRNGHALFSGFVEAAKH 538
DB 483 KAGIKVSGSLADKLVEMTENPHVFAAOFBEPSTPRDGHPLFSGFVEAAKH 539

```

## RESULT 8

ID PYRG VIBCH STANDARD; PRT; 544 AA.

```

AC 09KPC4;
DT 28-FEB-2003 (Rel. 41, Created)
DT 28-FEB-2003 (Rel. 41, Last sequence update)
DT 28-FEB-2003 (Rel. 41, Last annotation update)
DB CTP synthase (EC 6.3.4.2) (UTP--ammonia ligase) (CTP synthetase).
GN PYRG OR VC2448.
OS Vibrio cholerae.
OC Bacteria; Proteobacteria; Gammaproteobacteria; Vibrionales;
Vibrionaceae; Vibrrio.
OX NCBI_TaxID=666;
RN RP SEQUENCE FROM N.A.
RA STRAIN=El Tor N16961 / Serotype O1;
RX MEDLINE=20406833; PubMed=10952301;
RA Heidelberg J.F., Eisen J.A., Nelson W.C., Clayton R.A., Gwinn M.L.,

```



QY 2 TKF1PITGVVSLKGIKIAASSLAALIEDRGKVTITLDPYINVDPGTMSPPQHGVEV 61  
 DB 3 TNYIFVTOGVVSLKGIKIAASSLAALIEDRGKVTITLDPYINVDPGTMSPPQHGVEV 62  
 QY 62 TEDGERTDLDGHERELKMTKKNFTGOVYEQVIRNRKDYCATOVYPHITDE 121  
 DB 63 TQDGEATDLDGHERELKMTKKNFTGOVYEQVIRNRKDYCATOVYPHITDE 122  
 QY 122 IKRRTVSAEGKVALIEVGTVEDISLPFLFTIRQWVEAGSDRALFTHLTVPIKS 181  
 DB 123 IKSRVTDGAAGHDVAIVEGVTGVEDISLPFLFTIRQWVEAGSDRALFTHLTVPIKS 182  
 QY 182 AGEKTKRQTSVVELRTITGTPDLILRSRQPPASRRKIALFTVAEKAIVSAIDAD 241  
 DB 183 AGEVTKRQTSVVELRTITGTPDLILRSRQPPASRRKIALFTVAEKAIVSAIDAD 242  
 QY 242 TIVRPLLRQGLDVLVDQRLDVPADISAMEKVVDGLTHTDEVSAIVKRYVDHT 301  
 DB 243 STVOIPALKKQSGDLPICQRFHLDGPRADISEMGOVYQYDANPTGVEVIGWCKITELP 302  
 QY 302 DAYVSLNALIHAGITHHAKVQIYSTIDSETIABSTAYLKNVDAIIVPGGGERGVBSKI 361  
 DB 303 DAYVSVNALIHAGITHHAKVQIYSTIDSETIABSTAYLKNVDAIIVPGGGERGVBSKI 362  
 QY 362 STVRPARENKIPLYIGICGMSAVIEFARNVVGLEGASTFELPSPHPIVGLITENWDE 421  
 DB 363 LTAKARKNEIPLYIGICGMSAVIEFARNVVGLEGASTFELPSPHPIVGLITENWDE 422  
 QY 422 AGEVTRDESDLDGTMELGAQCKLADSLAPOLYQDVITERRHRYEPNNQYLKQLE 481  
 DB 423 EGNIEETRDADLDGTMELGAQCKLADSLAPOLYQDVITERRHRYEPNNQYLKQLE 482  
 QY 482 AAGMKESKSLDGRVETIELPEHPMFLACQHPFTSTPENGHALFSGFEVA 535  
 DB 483 KAGLVATLSADKCLVELLEVPNHPFVACQHPFTSTPENGHALFSGFEVA 536  
 RESULT 10  
 ID PYRG ECOLI STANDARD; PRT: 544 AA.  
 AC P08358; 01-AUG-1988 (Rel. 08, Created)  
 DT 01-OCT-1996 (Rel. 34, Last sequence update)  
 DT 15-SEP-2003 (Rel. 42, Last annotation update)  
 DE CTP synthase (EC 6.3.4.2) (UTP--ammonia ligase) (CTP synthetase).  
 GN PYRG OR B2780 OR C3345 OR Z4095 OR ECS3640 OR SF2795.  
 OS Escherichia coli  
 OS Escherichia coli O6,  
 OS Escherichia coli O157:H7, and  
 OS Shigella flexneri.  
 OC Bacteria; Proteobacteria; Gammaproteobacteria; Enterobacteriales;  
 OC Enterobacteriaceae; Escherichia.  
 OX NCBI\_TaxID=562, 217992, 83334, 623;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC SPECIES=E.coli;  
 RX MEDLINE=86168304; PubMed=3514618;  
 RA Weng M., Makaroff C.A., Zaikin H.;  
 RT "Nucleotide sequence of Escherichia coli pyrg encoding CTP  
 RT synthetase.";  
 RL J. Biol. Chem. 261:5568-5574 (1986).  
 RN [2]  
 RP SEQUENCE FROM N.A.  
 RC SPECIES=E.coli; STRAIN=K12 / MG1655;  
 RX MEDLINE=97426617; PubMed=9278503;  
 RA Blattner F.R., Plunkett G. III, Bloch C.A., Perna N.T., Burland V.,  
 RA Riley M., Collado-Vides J., Glasner J.D., Rode C.K., Mayhew G.F.,  
 RA Gregor J., Davis N.W., Kirkpatrick H.A., Goeden M.A., Rose D.J.,  
 RA Mau B., Shao Y.;  
 RT "The complete genome sequence of Escherichia coli K-12.";  
 RL Science 277:1453-1474 (1997).  
 RN [3]

RP SEQUENCE FROM N.A.  
 RC SPECIES=E.coli; STRAIN=O6:H1 / CFT073 / ATCC 700928;  
 RX MEDLINE=22388234; PubMed=12471157;  
 RA Welch R.A., Burland V., Plunkett G. III, Redford P., Roesch P.,  
 RA Rasko D., Buckles E.L., Zhou S.-R., Boutin A., Hackett J., Stroud D.,  
 RA Mayhew G.F., Rose D.J., Hsu S., Schwartz D.C., Perna N.T.,  
 RA Mobley H.L.T., Donnenberg M.S., Blattner F.R.;  
 RT "Extensive mosaic structure revealed by the complete genome sequence  
 RT of uropathogenic Escherichia coli.";  
 RL Proc. Natl. Acad. Sci. U.S.A. 99:17020-17024 (2002).  
 RN [4]  
 RP SEQUENCE FROM N.A.  
 RC SPECIES=E.coli; STRAIN=O157:H7 / EDL933 / ATCC 700927;  
 RX MEDLINE=21074935; PubMed=11206551;  
 RA Perna N.T., Plunkett G. III, Burland V., Mau B., Glasner J.D.,  
 RA Rose D.J., Mayhew G.F., Evans P.S., Gregor J., Kirkpatrick H.A.,  
 RA Postal G., Hackett J., Klink S., Boutin A., Shao Y., Miller L.,  
 RA Grobbeck E.J., Davis N.W., Lim A., Dimalanta E.T., Potomousis K.,  
 RA Apodaca J., Anantharaman T.S., Lin J., Yen G., Schwartz D.C.,  
 RA Welch R.A., Blattner F.R.;  
 RT "Genome sequence of enterohaemorrhagic Escherichia coli O157:H7.";  
 RL Nature 409:529-533 (2001).  
 RN [5]  
 RP SEQUENCE FROM N.A.  
 RC SPECIES=E.coli; STRAIN=O157:H7 / RIMD 0509952;  
 RX MEDLINE=21156231; PubMed=11258796;  
 RA Hayashi T., Makino K., Onishi M., Kurokawa K., Ishii K., Yokoyama K.,  
 RA Han C.-G., Ohtsubo E., Nakayama K., Murata T., Tanaka M., Tobe T.,  
 RA Kida S., Takami H., Honda T., Sasaki K., Ogasawara N., Yasunaga T.,  
 RA Ikura S., Shiba T., Hattori M., Shinagawa H.;  
 RT "Complete genome sequence of enterohaemorrhagic Escherichia coli  
 RT O157:H7 and genomic comparison with a laboratory strain K-12.";  
 RL DNA Res. 8:11-22 (2001).  
 RN [6]  
 RP SEQUENCE OF 1-12.  
 RC SPECIES=E.coli; STRAIN=K12 / EMG2;  
 RX MEDLINE=97443975; PubMed=9298646;  
 RA Link A.U., Robinson K., Church G.M.;  
 RT "Comparing the predicted and observed properties of proteins encoded  
 RT in the genome of Escherichia coli K-12.";  
 RL Electrophoresis 18:1259-1313 (1997).  
 RN [7]  
 RP SEQUENCE FROM N.A.  
 RC SPECIES=S.flexneri; STRAIN=301 / Serotype 2a;  
 RX MEDLINE=22272406; PubMed=123845590;  
 RA Jin Q., Yuan Z., Xu J., Wang Y., Shen Y., Lu W., Wang J., Liu H.,  
 RA Yang J., Yang F., Zhang X., Zhang J., Yang G., Hu H., Qu D., Dong J.,  
 RA Sun L., Xue Y., Zhao A., Gao Y., Zhu J., Kan B., Ding K., Chen S.,  
 RA Cheng H., Yao Z., He B., Chen R., Ma D., Qiang B., Wen Y., Hou Y.,  
 RA Yu J.;  
 RT "Genome sequence of Shigella flexneri 2a: insights into pathogenicity  
 RT through comparison with genomes of Escherichia coli K12 and O157.";  
 RL Nucleic Acids Res. 30:4432-4441 (2002).  
 RN [8]  
 RP MOTAGENESIS OF VAL-346; GLY-350 AND GLY-351.  
 RC SPECIES=E.coli;  
 RX MEDLINE=87250264; PubMed=3298209;  
 RA Weng M., Zaikin H.;  
 RT "Structural role for a conserved region in the CTP synthetase  
 RT glutamine amide transfer domain.";  
 RL J. Bacteriol. 169:3023-3028 (1987).  
 RN [9]  
 RP ENZYME REGULATORS.  
 RC SPECIES=E.coli;  
 RX MEDLINE=3322179; PubMed=8385490;  
 RA Robertson U.G., Villafraña J.U.;  
 RT "Characterization of metal ion activation and inhibition of CTP  
 RT synthetase.";  
 RL Biochemistry 32:3769-3777 (1993).  
 RN [10]  
 RP ROLE OF GTP, AND MOTAGENESIS OF CYS-378.  
 RC SPECIES=E.coli;  
 RX MEDLINE=21234865; PubMed=11336655;





Query Match 68.5%; Score 1908.5; DB 1; Length 545;  
 Best Local Similarity 67.5%; Pred. No. 3e-112;  
 Matches 364; Conservative 71; Mismatches 103; Indels 1; Gaps 1;

2 TXRFITFGVSSISGKGIASSIAALIEDRGKATITTKIDPYINVDGMSPPGHGEV 61  
 3 TNYIFETGGVSSISGKGIASSIAALIEDRGKATITTKIDPYINVDGMSPPGHGEV 62  
 62 TEDGATDLDLGHYERFLKTTMKNNFTTGOVEOLNRERKDYLGATVOVPHITDE 121  
 63 TCGAETDLDLGHYERFLKTTMKNNFTTGOVEOLNRERKDYLGATVOVPHITDE 122  
 122 IKRRVYASAGKQVALIEVGCTGDIESTPLETINOMGVELGRDALFIHLTPVYKS 181  
 123 IKORVIAQGHQHVIVVEGVGTGDIESTPLETINOMGVELGRDALFIHLTPVYKS 182  
 182 AGELKTPQHSVKEKRTIGIQPDILICSEOPIPASERKIALFTNVAEKAVISADAD 241  
 183 AGEVKTQPHQSVKELSLIGIQPDVLCISDPMIPNERAKIALFCNVAERAVISADAD 242  
 242 TIYRIFLLRBOGLDVLVDQLRLDVPADLSAMEKYVDGLTHPTDEVSIALVGYDHT 301  
 243 SIYQIFALKKSGQGLDPEVCEFRFLTCPEADLTMEGVLYKQANPVEVITIGWVXTLEP 302  
 302 DAKSINELALIHAGIHRHVKVOISYIDSEITBAGFAKIKNDALIVPGSPGREGKI 361  
 303 DAKSVNELKAKAGLNRSLSVNIKTYIDSOVETKGVYELKIDGILVPGSPGREGKI 362  
 362 STYVFARENKIPYLGICLQMSQAVIEFPAANVGLRGAHSTELPSPKSPHVLIGLITEMWD 421  
 363 RPAQVARENKIPYLGICLQMSQAVIEFPAANVGLRGAHSTELPSPKSPHVLIGLITEMWD 422  
 422 AGEVLTRDSDSLGCTMRIGAKCRKADSLAQLQKOVITERRHREPNNOYAKOL 481  
 423 EGNTVEVTRSDSLGCTMRIGAKCRKADSLAQLQKOVITERRHREPNNOYAKOL 482  
 482 AAGMKFSKSLDGLRLVEIIELEPHFPLACQFHEFTSTPRNHALSGFVENA-AKHK 539  
 483 KAGLKYTGSLADKKLVEIIEVPHFPAVACQFHEFTSTPRNHALSGFVENA-AKHK 541

RESULT 13  
 PYRG\_XYLF 1 STANDARD; PRT; 554 AA.

AC 09PDTL;  
 DT 28-FEB-2003 (Rel. 41, Last sequence update)  
 DT 28-FEB-2003 (Rel. 41, Last annotation update)  
 DE CTP synthase (EC 6.3.4.2) (UTP--ammonia ligase) (CTP synthetase).  
 GN PYRG OR XF1288.  
 OS Xylella fastidiosa.  
 OC Bacteria; Proteobacteria; Gammaproteobacteria; Xanthomonadales;  
 OC Xanthomonadaceae; Xylella.  
 NCBI\_TaxID=2371;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=9a5c;  
 RX MEDLINE=20365717; PubMed=10910347;  
 RA Alvarado A.J.G., Reinach F.C., Arruda P., Abreu F.A., Acencio M.,  
 RA Alvarado A.J.G., Alves L.M.C., Araya J.E., Baia G.S., Baptista C.S.,  
 RA Barros M.H., Bonaccorsi E.D., Bordin S., Bove J.M., Britona M.R.S.,  
 RA Bueno M.R.P., Camargo A.A., Camargo L.E.A., Carraro D.M., Carier H.,  
 RA Colatto N.B., Colombo C., Costa F.F., Costa M.C.R., Costa-Neto C.M.,  
 RA Coutinho L.L., Cristofani M., Dias-Neto E., Docena C., El-Dorry H.,  
 RA Facincani A.P., Ferreira A.J.S., Ferreira V.C.A., Perro J.A.,  
 RA Faria J.S., Frasca S.C., Goldman M.C., Frohne M., Furlan L.R.,  
 RA Garner L., Gindan G.H., Goldman M.H.S., Gomes S.L., Gruber A.,  
 RA Ho P.L., Honneisel J.D., Junqueira M.L., Kemper B.L., Katsjima J.P.,  
 RA Krieger J.E., Kuzmae B.E., Laigret F., Lambais M.R., Leite L.C.C.,  
 RA Lemos E.G.M., Lemos M.V.F., Lopes S.A., Lombardi C.R., Machado J.A.,  
 RA Machado M.A., Madeira A.M.B.N., Madeira H.M.F., Marino C.L.,  
 RA Marques M.V., Martins E.A.L., Martins S.M.F., Matsukuma A.Y.,  
 RA Menck C.F.M., Mitraca E.C., Miyaki C.Y., Monteiro-Vitorello C.B.,

RA Moon D.H., Nagai M.A., Nascimento A.L.T.O., Netto L.E.S.,  
 RA Nhami A., Jr., Nobrega F.G., Nunes L.R., Oliveira M.A.,  
 RA de Oliveira M.C., de Oliveira R.C., Palmeri D.A., Paris A.,  
 RA Peixoto B.R., Pereira G.A.G., Pereira H.A.Jr., Pasquero J.B.,  
 RA Quaggio R.B., Roberto P.G., Rodrigues V., de Rosa A.J.M.,  
 RA de Rosa V.E., Jr., de Sa R.G., Sancheli R.V., Sawaaki H.E.,  
 RA da Silva A.C.R., da Silva A.M., da Silva F.R., Silva M.A.Jr.,  
 RA da Silveira J.P., Silvestri M.L.Z., Siqueira W.J., de Souza A.A.,  
 RA de Souza A.P., Terenzi M.F., Truffi D., Tsai S.M., Tsubako M.H.,  
 RA Vallada H., Van Sluys M.A., Verjovski-Almeida S., Vettore A.L.,  
 RA Zago M.A., Zatz M., Zeldman J., Zetser J.C.,  
 RT The genome sequence of the plant pathogen Xylella fastidiosa.  
 CC Nature 406:151-159(2000).  
 CC -FUNCTION: Catalyzes the ATP-dependent amination of UTP to CTP with  
 CC either L-glutamine or ammonia as the source of nitrogen (By  
 CC similarity).  
 CC -CATALYTIC ACTIVITY: ATP + UTP + NH(3) = ADP + phosphate + CTP.  
 CC -ENZYME REGULATION: Allosterically activated by GTP, when glutamine  
 CC is the substrate. Inhibited by CTP (By similarity).  
 CC -PATHWAY: Pyrimidine biosynthesis; conversion of UMP to CTP; third  
 CC (last) step.  
 CC -SUBUNIT: Homotrimer (By similarity).  
 CC -SIMILARITY: Belongs to the CTP synthase family.  
 CC -SIMILARITY: Contains 1 type-1 glutamine amidotransferase domain.  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
 CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).

DR EMBL: AE003962; AAF64097.1; -  
 DR PIR: C82700; C82700.  
 DR HAMAP: MF\_01227; -; 1.  
 DR InterPro: IPR000991; GATase\_1.  
 DR InterPro: IPR004468; PYRG.  
 DR Pfam: PF00117; GATase: 1.  
 DR TIGRFAMs: TIGR00337; PYRG: 1.  
 DR PROSITE: PS00442; GATase\_type\_1; 1.  
 DR Pyrimidine biosynthesis; Ligase; Glutamine amidotransferase;  
 KW Complete proteome.  
 FT DOMAIN 1 301 AMINATOR DOMAIN.  
 FT 302 554  
 FT ACT SITE 380 380 GATASE (BY SIMILARITY).  
 FT ACT SITE 518 518 GATASE (BY SIMILARITY).  
 FT ACT SITE 520 520 GATASE (BY SIMILARITY).  
 SQ SEQUENCE 554 AA; 61558 MW; 5EB8769917043811 CRC64;

Query Match 68.3%; Score 1905; DB 1; Length 554;  
 Best Local Similarity 66.1%; Pred. No. 5.1e-112;  
 Matches 361; Conservative 79; Mismatches 102; Indels 4; Gaps 2;

1 MTKRIFITGGVSSISGKGIASSIAALIEDRGKATITTKIDPYINVDGMSPPGHGEV 60  
 1 MTKRIFITGGVSSISGKGIASSIAALIEDRGKATITTKIDPYINVDGMSPPGHGEV 60  
 61 VTDGATDLDLGHYERFLKTTMKNNFTTGOVEOLNRERKDYLGATVOVPHITD 120  
 61 VTDGATDLDLGHYERFLKTTMKNNFTTGOVEOLNRERKDYLGATVOVPHITD 120  
 121 IKRRVYASAGKQVALIEVGCTGDIESTPLETINOMGVELGRDALFIHLTPVYKS 180  
 121 IKRRVYASAGKQVALIEVGCTGDIESTPLETINOMGVELGRDALFIHLTPVYKS 180  
 122 IKORVIAQGHQHVIVVEGVGTGDIESTPLETINOMGVELGRDALFIHLTPVYKS 180  
 182 AGELKTPQHSVKEKRTIGIQPDILICSEOPIPASERKIALFTNVAEKAVISADAD 240  
 183 AGEVKTQPHQSVKELSLIGIQPDVLCISDPMIPNERAKIALFCNVAERAVISADAD 240  
 241 TIYRIFLLRBOGLDVLVDQLRLDVPADLSAMEKYVDGLTHPTDEVSIALVGYDHT 298  
 241 TIYRIFLLRBOGLDVLVDQLRLDVPADLSAMEKYVDGLTHPTDEVSIALVGYDHT 298  
 242 SIYQIFALKKSGQGLDPEVCEFRFLTCPEADLTMEGVLYKQANPVEVITIGWVXTLEP 300

QY 299 DHTDAKSLNEALIHAGHTRHKVQISYDSETEAEATKAKNDALIVPGSGERGVE 358  
 DB 301 DYQADAKSVGAEALKHGKGLKORTKVMKWEADLESDMGALKDIDGLIVPGSGFDGRPE 360  
 QY 359 GKISTVPRPARENKIPYIGICLQKQSAVIEFARNVVGLEGASHTEFLPKSPHVVIGLITWM 418  
 DB 361 GKVALSRVAREQVRVPEFGICGMQAAVVOYARHVAAGLEAGANSTENDROSPHVVIALITWM 420  
 QY 419 MDEAGELVTRDESDSDIGTWRLGAQKCRUKADSLAFOLYQKDVITERHRRHRYEPNNQYIK 478  
 DB 421 RITTCGEVERDEKSDIGTWRLGAQKCRUKADSLAFOLYQKDVITERHRRHRYEPNNQYIK 480  
 QY 479 QLEAAKTKFSKSLDRIYVEIIELEP--HPWFLACQFHEPFTSTPRNGHALPSGVEEAA 536  
 DB 481 QLEDAELVTAAKSDDTLVEMLBLPRNHPWFLACQFHEPFTSTPRNGHALPSGVEEAA 540  
 QY 537 KHKTOG 542  
 DB 541 ARKAGG 546

## RESULT 14

ID PYRG\_NEIMA STANDARD; PRT; 544 AA.

AC 09UTL;  
 DT 28-FEB-2003 (Rel. 41, Last sequence update)  
 DT 28-FEB-2003 (Rel. 41, Last annotation update)  
 DE CTP synthase (EC 6.3.4.2) (UTP--ammonia ligase) (CTP synthetase).  
 GN PYRG OR NMB154.  
 OS Neisseria meningitidis (serogroup A).  
 OC Bacteria; Proteobacteria; Betaproteobacteria; Neisseriales;  
 OC Neisseriaceae; Neisseria.  
 OX NCBI\_TaxID=65689;  
 RN (1)  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=Z2491 / Serotype 4A;  
 RX MEDLINE=20222556; PubMed=10761919;  
 RA Parkhill J., Achtman M., James K.D., Bentley S.D., Churcher C.,  
 Klee S.R., Morelli G., Basham D., Brown D., Chillingworth T.,  
 Davies R.M., Davis P., Devlin K., Felwell T., Hamlin N., Holtroyd S.,  
 Jagers K., Leach S., Moule S., Mungall K., Quail M.A.,  
 Rajandream M.A., Rutherford K.M., Simmonds M., Skelton J.,  
 Whitehead S., Spratt B.G., Barrell B.G.;  
 RA "Complete DNA sequence of a serogroup A strain of Neisseria  
 meningitidis Z2491."  
 RT Nature 404:502-506 (2000).  
 RL -1- FUNCTION: Catalyzes the ATP-dependent amination of UTP to CTP with  
 either L-glutamine or ammonia as the source of nitrogen (by  
 similarity).  
 CC -1- CATALYTIC ACTIVITY: ATP + UTP + NH(3) = ADP + phosphate + CTP.  
 CC -1- ENZYME REGULATION: Allosterically activated by GTP, when glutamine  
 is the substrate. Inhibited by CTP (by similarity).  
 CC -1- PATHWAY: Pyrimidine biosynthesis; conversion of UMP to CTP; third  
 (last) step.  
 CC -1- SUBUNIT: Homotrimer (by similarity).  
 CC -1- SIMILARITY: Belongs to the CTP synthase family.  
 CC -1- SIMILARITY: Contains 1 type-1 glutamine amidotransferase domain.  
 CC This Swiss-Prot entry is copyright. It is produced through a collaboration  
 between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 the European Bioinformatics Institute. There are no restrictions on its  
 use by non-profit institutions as long as its content is in no way  
 modified and this statement is not removed. Usage by and for commercial  
 entities requires a license agreement (See <http://www.isd-sib.ch/announce/>  
 or send an email to [license@sib-sib.ch](mailto:license@sib-sib.ch)).  
 CC EMBL: AL162757; CAB84970.1;  
 DR PIR: F01798; P01798.  
 DR HAMAP: MF\_01227; 1.  
 DR InterPro: IPR000991; GATase\_1.  
 DR InterPro: IPR004468; Pyrg.

DR Pfam: PF00117; GATase; 1.  
 DR TIGRfam: TIGR00337; Pyrg; 1.  
 DR PROSITE: PS00442; GATase type 1; 1.  
 KW Pyrimidine biosynthesis; ligase; Glutamine amidotransferase;  
 KM Complete proteome.  
 FT DOMAIN 1 299 AMINATOR DOMAIN.  
 FT ACT\_SITE 300 544 GLUTAMINE AMIDOTRANSFERASE.  
 FT ACT\_SITE 380 380 GATASE (BY SIMILARITY).  
 FT ACT\_SITE 517 517 GATASE (BY SIMILARITY).  
 FT ACT\_SITE 519 519 GATASE (BY SIMILARITY).  
 SQ SEQUENCE 544 AA; 59916 MW; 99376E7C6557CAE2 CRC64;

Query Match 67.9%; Score 1892.5; DB 1; Length 544;  
 Best Local Similarity 66.1%; Pred. No. 3e-111;  
 Matches 358; Conservative 80; Mismatches 10; Indels 3; Gaps 2;

QY 1 MTKRIFINGVAVSSILGKGIASLAAILBDRGLRVITTKLDPIYINVDPGTWSPFQHGVEF 60  
 DB 1 MTKRIFVGVAVSSILGKGIASLAAILBDRGLRVITTKLDPIYINVDPGTWSPFQHGVEF 60  
 QY 61 VTDEGAEFTDLGHYERFLTKTMTKKNFTGOYVEQVLRNRRGQVLAGATVOYIPIHTD 120  
 DB 61 VTDEGAEFTDLGHYERFLTKTMTKKNFTGOYVEQVLRNRRGQVLAGATVOYIPIHTD 120  
 QY 121 EIKRVESAEKGQVAVLIEVGTVGDIESTPFEITIRQMGVGLGRDALFTHTLVPIYK 180  
 DB 121 EIKRVHGAQGVAVLIEVGTVGDIESTPFEITIRQMGVGLGRDALFTHTLVPIYK 180  
 QY 181 SAGELKTPYTHSYKELTGTIGDPIICSEQETIPASERKIALFTYNAKAVISALDA 240  
 DB 181 AAGRIKXKPTHTYKEMLSIGLQDPIICRDRTPMPADERKIALFCVERRAIVGSYDV 240  
 QY 241 DTIRIFPLLRQGLDLDVQDLADVPADLSAMEKVDLTHPTDEVSIAIVGYVDH 300  
 DB 241 DSIYCEPMHLDQSDIDNITITQLOLVQADLTIMKCIYALIQPKTVKIAMGVYDL 300  
 QY 301 TDVAKSLNEALIHAGHTRHKVQISYDSETEAEG--TAKLKNVDALIVPGSGERGVE 358  
 DB 301 TDSYKSLLEALKHAGVTEITVQITPVDSEIERKNGQVSLKMDALIVPGSGERGVE 360  
 QY 359 GKISTVPRPARENKIPYIGICLQKQSAVIEFARNVVGLEGASHTEFLPKSPHVVIGLITWM 418  
 DB 361 GKVALSRVAREQVRVPEFGICGMQAAVVOYARHVAAGLEAGANSTENDROSPHVVIALIDEM 420  
 QY 419 MDEAGELVTRDESDSDIGTWRLGAQKCRUKADSLAFOLYQKDVITERHRRHRYEPNNQYIK 478  
 DB 421 QTADGVSATREDSEADIGTWRLGAQKCRUKADSLAFOLYQKDVITERHRRHRYEPNNQYIK 480  
 QY 479 QLEAAKTKFSKSLD-GRIVEIIELEP--HPWFLACQFHEPFTSTPRNGHALPSGVEEAAK 537  
 DB 481 TLEQAGLVIGVSGARELVETIELPNHPWFLACQFHEPFTSTPRNGHALPSGVEEAAK 540  
 QY 538 HK 539  
 DB 541 NK 542

## RESULT 15

ID PYRG\_NEIMA STANDARD; PRT; 544 AA.

AC 09UTL;  
 DT 28-FEB-2003 (Rel. 41, Last sequence update)  
 DT 28-FEB-2003 (Rel. 41, Last annotation update)  
 DE CTP synthase (EC 6.3.4.2) (UTP--ammonia ligase) (CTP synthetase).  
 GN PYRG OR NMB154.  
 OS Neisseria meningitidis (serogroup B).  
 OC Bacteria; Proteobacteria; Betaproteobacteria; Neisseriales;  
 OC Neisseriaceae; Neisseria.  
 OX NCBI\_TaxID=491;  
 RN (1)  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=MC58 / Serogroup B;

MEBLINE=20175755; PubMed=10710307;  
RA Terebin H., Saunders N.J., Heideberg J., Jeffries A.C., Nelson K.E.,  
RA Eisen J.A., Kechum R.A., Hoog D.W., Peden J.F., Dodson R.J.,  
RA Nelson W.C., Gwin M.L., Deboy R., Peterson J.D., Hickey E.K.,  
RA Hatt D.H., Salzberg S.L., White O., Fleischmann R.D., Dougherty B.A.,  
RA Mason T., Ciecko A., Parksey D.S., Blair E., Ciltone H., Clark E.B.,  
RA Cotton M.D., Utterback T.R., Khouri H., Qin H., Vannotevan J.,  
RA Gill J., Scariato V., Masigant V., Pizzo M., Grandi G., Sun L.,  
RA Smith H.O., Fraser C.M., Moxon E.R., Rappelli R., Venter J.C.,  
RA "Complete genome sequence of *Neisseria meningitidis* serogroup B strain  
MC58.";  
RT Science 287:1809-1815 (2000).  
CC -1- FUNCTION: Catalyzes the ATP-dependent amination of UTP to CTP with  
CC either L-glutamine or ammonia as the source of nitrogen [by  
CC similarity].  
CC -1- CATALYTIC ACTIVITY: ATP + UTP + NH(3) = ADP + phosphate + CTP.  
CC -1- ENZYME REGULATION: Allosterically activated by GTP, when glutamine  
CC is the substrate. Inhibited by CTP (by similarity).  
CC -1- PATHWAY: Pyrimidine biosynthesis; conversion of UMP to CTP; third  
CC (last) step.  
CC -1- SUBUNIT: Homotetramer (by similarity).  
CC -1- SIMILARITY: Contains 1 type-1 glutamine amidotransferase domain.  
CC -----  
CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
CC the European Bioinformatics Institute. There are no restrictions on its  
CC use by non-profit institutions as long as its content is in no way  
CC modified and this statement is not removed. Usage by, and for commercial  
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
CC -----  
DR EMBL: AE002505; AAF41908.1; -  
DR PIR: F81070; F81070.  
DR TIGR: NMB1554; -  
DR HAMAP: MF\_01227; -; 1.  
DR InterPro: IPR000991; GATase\_1.  
DR InterPro: IPR004468; Pyrg.  
DR Pfam: PF00117; GATase; 1.  
DR TIGRFAMs: TIGR00337; Pyrg; 1.  
DR PROSITE: PS00442; GATase TYPE 1; 1.  
DR Pyrimidine biosynthesis; Llgase; Glutamine amidotransferase;  
KW Complete proteome.  
FT DOMAIN 1 299 AMINATOR DOMAIN.  
FT DOMAIN 300 544 GLUTAMINE AMIDOTRANSFERASE.  
FT ACT SITE 380 380 GATASE (BY SIMILARITY).  
FT ACT SITE 517 517 GATASE (BY SIMILARITY).  
FT ACT SITE 519 519 GATASE (BY SIMILARITY).  
SQ SEQUENCE 544 AA; 59923 MW; E3C67621DF52B69C CRC64;

Query Match 67.7%; Score 1888.5; DB 1; Length 544;  
Best Local Similarity 66.2%; Pred. No. 5.4e-111;  
Matches 359; Conservative 77; Mismatches 103; Indels 3; Gaps 2;

QY 1 MKKFFINGVSSSGKGIASSLAILEDKGLKVTITKLDPIYNDPGTMSPROHGEVP 60  
DB 1 MKKFLVTVGVSSSGKGIASSLAILEDKGLKVTITKLDPIYNDPGTMSPROHGEVP 60  
QY 61 VTEDGAETDLDLGHYERFLKTTMTKKNFTTQVYEQVLAENERKDYAGATVQVPIPHITD 126  
DB 61 VTDDGAETDLDLGHYERFLKTTMTKKNFTTQVYEQVLAENERKDYAGATVQVPIPHITD 126  
QY 121 EIKRRVVSAGCKVALIEVGCTGDIISLFLFTITROMGVELGRDALPIHLTLVPYIK 180  
DB 121 EIKRRVVSAGCKVALIEVGCTGDIISLFLFTITROMGVELGRDALPIHLTLVPYIA 180  
QY 181 SAGELKTPQSHVYELTITIGDPIILCRSEOPIPASERKIALFTNVAEKAVISAIDA 240  
DB 181 AAGIKTKPQSHVYELTITIGDPIILCRDRMPADERRKIALFCNVERAIVGSYDV 240  
QY 241 DTIYRIPLLREOGDLVDQLADVPADLSAMEKYVDGLTEPTDEVSAIVGKTVDH 300  
DB 241 DSIYCEPEMLHDQGDINITEQLQANVOQADLTAMKKIVHAIQNPKHIVKIAMVGKYVDL 300

QY 301 TDAYSINAEALIHAGIHTRHKVOISYDSETEAE--GTAKLKNVDALIVGFGGSGRVE 358  
DB 301 TBSYKSLIEALKHAGIHTETDVOITFPDSENIENKPKDVSMLKMDAILVGGGSGRVE 360  
QY 359 GKISTVAPARENKIPYIGICGMSQAYIEFARNVVGEGASTFELKSPHPVIGLITW 418  
DB 361 GKIAVRYARENVPYIGICGMSQAYIEFARNVVGEGASTFELKSPHPVIGLITW 420  
QY 419 MDEAGELVTRDESDLDGIVWLAGQKRLKADSLAPOLYOKDVITERRRRERFENNQYLK 478  
DB 421 QTAGDSVETRESDLDGIVWLAGQVEVLKXGSAAAYIGSGHIRERRRERFENNQYVP 480  
QY 479 QLEAAKGFSGSKLD-GRVPIITLPEHPWFLAQCFPEPSTFRNGHALFSGVPEAAK 537  
DB 481 TLEQAGLVIGVSAGREBLVETIELPNHPWFFACQFPEFTSPRKGPLPTAFVKAALN 540  
QY 538 HK 539  
DB 541 NK 542

Search completed: January 29, 2004, 15:50:42  
Job time : 11.244 secs



```

QY 182 AGELTKPTQSHSVKELRTTIGIOPDILICRSEOPIPASERKIALFTYVAKVISAIDND 241
DB 183 AGEVTKPTQSHSVKELRTTIGIOPDILICRSDRIPANERKIALFTYVAKVISAIDND 242
QY 242 TYIRPILRLRBOGLDLDVVDQLRDVPAADLSAMERYVDGLTHPTDEVSIALVGYVDHT 301
DB 243 SYIKPIQLRSQGLDLDVVKRFSITCREADLSSEWENVIYOGANRNGEVLIMVCKIYELP 302
QY 302 DAYSLNEMALIHAGIHRHKVQISYIDSETEBAAGTAKLVNDAILVPGGGEVGECKI 361
DB 303 DAYSVNEMALIHAGIHRHKVQISYIDSETEBAAGTAKLVNDAILVPGGGEVGECKI 362
QY 362 STYVFARENKIPYIGICIGMOSAVIEFARNVVGLEGASTPELPKSPHPVIGLITEMMD 421
DB 363 LAAYFARENKIPYIGICIGMOSAVIEFARNVVGLEGASTPELPKSPHPVIGLITEMMD 422
QY 422 AGEVLTRDESDLDIGTMRIGAQKCRLLKADSLAFQLYQKDVITERHRRHRYEFNNQYLKQLE 481
DB 423 EGVNVEOREHESADLDIGTMRIGAQKCRLLKADSLAFQLYQKDVITERHRRHRYEFNNQYLKQLE 482
QY 482 AAGMKESGKSLDGRVLEIIELEPHPMFLACQFHEPSTTPRNGHALFSGFEVAAAKH 538
DB 483 QAGLVFSGLSDRKLVEMTLEKDPHPVAGQFHEPSTTPRNGHALFSGFEVAAAKH 539

```

## RESULT 2

```

QY 08EB09 PRELIMINARY; PRT; 546 AA.
AC 08EB09:
DT 01-MAR-2003 (TREMELrel. 23, Created)
DT 01-MAR-2003 (TREMELrel. 23, Last sequence update)
DE CTP synthase.
OS Pyrog. OR SO3441.
OC Shewanella oneidensis.
OC Bacteria; Proteobacteria; Gammaproteobacteria; Alteromonadales;
OC Alteromonadaceae; Shewanella.
OX NCBI_TaxID=70863;
RN (1)
RP SEQUENCE FROM N.A.
RX MEDLINE=22297666; PubMed=12368613;
RA Heidelberg J.F., Paulsen I.T., Nelson K.B., Gaidos E.J., Nelson W.C.,
RA Meyer T.D., Eisen J.A., Seshadri R., Ward N., Methe B., Clayton R.A.,
RA DeBoy R.T., Dodson R.J., Durkin A.S., Hatt D.H., Kolonay J.F.,
RA Madupu R., Peterson J.D., Umayam L.A., White O., Wolf A.M.,
RA Vamathevan J., Weidman J., Imprimis M., Lee K., Berry K., Lee C.,
RA Mueller J., Khouri H., Gill J., Uitterlinden T.R., McDonald L.A.,
RA Feldblyum T.V., Smith H.O., Venter J.C., Neilson K.H., Fraser C.M.;
RT "Genome sequence of the dissimilatory metal ion-reducing bacterium
RT Shewanella oneidensis."
RL Nat. Biotechnol. 20:1118-1123(2002).
DR EMBL; AB015780; AAN56438.1; -.
DR TIGR; SO3441; -.
KM Complete proteome.
SQ SEQUENCE 546 AA; 60148 MW; 4079A1DB93CF39E7 CRC64;

```

```

Query Match 69.0%; Score 1925; DB 16; Length 546;
Best Local Similarity 68.2%; Pred. No. 2e-125;
Matches 367; Conservative 69; Mismatches 102; Indels 0; Gaps 0;
QY 2 TKPIFTGGVSSLGKIAASSLAALIEDRGKVTITTKLDPIYNDPGTMSFGGEVAV 61
DB 3 TKPIFTGGVSSLGKIAASSLAALIEDRGKVTITTKLDPIYNDPGTMSFGGEVAV 62
QY 62 TEDGATDLDLGHYERELKTYTKNNFTTGQYEVGLANERKGVLGATVAVIPIHTD 121
DB 63 TEDGATDLDLGHYERELKTYTKNNFTTGQYEVGLANERKGVLGATVAVIPIHTD 122
QY 122 IKRRVYESAGDVALLIEVGTVGDIESTLPLETTIKQGVGLGRDALFIHLTLVPIYS 181
DB 123 IKRRVYESAGDVALLIEVGTVGDIESTLPLETTIKQGVGLGRDALFIHLTLVPIYS 182

```

```

DB 123 IKRRVYESAGDVALLIEVGTVGDIESTLPLETTIKQGVGLGRDALFIHLTLVPIYS 182
QY 182 AGELTKPTQSHSVKELRTTIGIOPDILICRSEOPIPASERKIALFTYVAKVISAIDND 241
DB 183 AGEVTKPTQSHSVKELRTTIGIOPDILICRSDRIPANERKIALFTYVAKVISAIDND 242
QY 242 TYIRPILRLRBOGLDLDVVDQLRDVPAADLSAMERYVDGLTHPTDEVSIALVGYVDHT 301
DB 243 SYIKPIQLRSQGLDLDVVKRFSITCREADLSSEWENVIYOGANRNGEVLIMVCKIYELP 302
QY 302 DAYSLNEMALIHAGIHRHKVQISYIDSETEBAAGTAKLVNDAILVPGGGEVGECKI 361
DB 303 DAYSVNEMALIHAGIHRHKVQISYIDSETEBAAGTAKLVNDAILVPGGGEVGECKI 362
QY 362 STYVFARENKIPYIGICIGMOSAVIEFARNVVGLEGASTPELPKSPHPVIGLITEMMD 421
DB 363 LAAYFARENKIPYIGICIGMOSAVIEFARNVVGLEGASTPELPKSPHPVIGLITEMMD 422
QY 422 AGEVLTRDESDLDIGTMRIGAQKCRLLKADSLAFQLYQKDVITERHRRHRYEFNNQYLKQLE 481
DB 423 EGVNVEOREHESADLDIGTMRIGAQKCRLLKADSLAFQLYQKDVITERHRRHRYEFNNQYLKQLE 482
QY 482 AAGMKESGKSLDGRVLEIIELEPHPMFLACQFHEPSTTPRNGHALFSGFEVAAAKH 539
DB 483 QAGLVFSGLSDRKLVEMTLEKDPHPVAGQFHEPSTTPRNGHALFSGFEVAAAKH 540

```

## RESULT 3

```

QY 08GB64 PRELIMINARY; PRT; 498 AA.
AC 08GB64:
DT 01-MAR-2003 (TREMELrel. 23, Created)
DT 01-MAR-2003 (TREMELrel. 23, Last sequence update)
DE CTP synthase (fragment).
OS Aeromonas hydrophila.
OC Bacteria; Proteobacteria; Gammaproteobacteria; Aeromonadales;
OC Aeromonadaceae; Aeromonas.
OX NCBI_TaxID=644;
RN (1)
RP SEQUENCE FROM N.A.
RA Sha J., Galindo C.L., Pancholi V., Popov V.L., Chopra A.K.;
RT "Identification of a potentially new virulence gene encoded that is
RT differentially expressed under in vitro versus in vivo growth
RT conditions in Aeromonas hydrophila."
RL Submitted (Aug-2002) to the EMBL/GenBank/DBJ databases.
DR EMBL; AY141757; AAN28925.1; -.
FT NON TER
SQ SEQUENCE 498 AA; 55079 MW; D9C6FC9228FBCB0 CRC64;

```

```

Query Match 62.3%; Score 1738; DB 2; Length 498;
Best Local Similarity 66.1%; Pred. No. 1.8e-112;
Matches 326; Conservative 70; Mismatches 97; Indels 0; Gaps 0;
QY 47 DPGTMSFGGEVAVTEEDGATDLDLGHYERELKTYTKNNFTTGQYEVGLANERKGV 106
DB 1 DPGTMSFGGEVAVTEEDGATDLDLGHYERELKTYTKNNFTTGQYEVGLANERKGV 107
QY 107 YLGATVQYIPIHTBIKRRVYESAGDVALLIEVGTVGDIESTLPLETTIKQGVGLGR 166
DB 108 YLGATVQYIPIHTBIKRRVYESAGDVALLIEVGTVGDIESTLPLETTIKQGVGLGR 167
QY 167 RALFTLTLVPIYSAGELTKPTQSHSVKELRTTIGIOPDILICRSEOPIPASERKIALF 226
DB 168 RALFTLTLVPIYSAGELTKPTQSHSVKELRTTIGIOPDILICRSDRIPANERKIALF 227
QY 227 TNYAEKAVISAIDNDLTIYRIPILRLRBOGLDLDVVDQLRDVPAADLSAMERYVDGLTHPT 286
DB 228 TNYAEKAVISAIDNDLTIYRIPILRLRBOGLDLDVVDQLRDVPAADLSAMERYVDGLTHPT 287
QY 287 DEVSAIYGVYVDVDAVYSALREALIHAGIHRHKVQISYIDSETEBAAGTAKLVNDAIL 346
DB 288 DEVSAIYGVYVDVDAVYSALREALIHAGIHRHKVQISYIDSETEBAAGTAKLVNDAIL 347
QY 347 AAVTIIGAVKVSIPDAVYSALREALIHAGIHRHKVQISYIDSETEBAAGTAKLVNDAIL 300
DB 348 AAVTIIGAVKVSIPDAVYSALREALIHAGIHRHKVQISYIDSETEBAAGTAKLVNDAIL 301

```

QY 347 LNRGCGEVEKISTVRARANKIPLYGICLQMSAVIEFARNVGLAGASTEPLEK 406  
 DB 301 LNVGGGESEVEKILAAQYARANKIPLYGICLQMSAVIEFARNVAGAGASSEPKD 360  
 QY 407 SPHPVIGLITENWDEAGELVTRDESDLGITWELGAOKCEKLSLAFOLYQCDVTEHH 466  
 DB 361 CAYPVGLITENWDDSGNVEITREKSDLGITWELGSLCHVDSKYRQMGSPITIEHH 420  
 QY 467 RHRVEYNNKILPQIEAAGLKVITGLSADKXIVELIIPDHFWVAQFHEPFIETPADGHA 480  
 DB 421 RHRVEYNNKILPQIEAAGLKVITGLSADKXIVELIIPDHFWVAQFHEPFIETPADGHA 480  
 QY 527 LFGSGVEAAAKH 539  
 DB 481 LFRAGFYKAGETQ 493

RESULT 4  
 ID 08D2K0 PRELIMINARY; PRT; 552 AA.  
 AC 08D2K0;  
 DT 01-MAR-2003 (TEMBLrel. 23, Created)  
 DT 01-MAR-2003 (TEMBLrel. 23, last sequence update)  
 DE 01-MAR-2003 (TEMBLrel. 23, last annotation update)  
 GN Pyrg protein.  
 OS Wigglesworthia brevipalpis.  
 OC Bacteria; Proteobacteria; Gammaproteobacteria; Enterobacteriales;  
 OC Enterobacteriaceae; Wigglesworthia.  
 NCBI\_TaxID=164609;  
 RX MEDLINE=22297718; PubMed=12219091;  
 RA Akman L., Yamashita A., Watanabe H., Oshima K., Shiba T., Hattori M.,  
 RT Genome sequence of the endocellular obligate symbiont of tsetse  
 RT files, Wigglesworthia glossinidia."  
 RL Nat. Genet. 32:402-407(2002).  
 DR EMBL; AB063522; BAC24500.1; -  
 KW Complete proteome.  
 SQ SEQUENCE 552 AA; 62336 MW; D7DC35A8FF91C568 CRC64;

Query Match 59.6%; Score 1662; DB 16; Length 552;  
 Best Local Similarity 58.7%; Pred. No. 4.1e-107;  
 Matches 320; Conservative 90; Mismatches 123; Indels 12; Gaps 2;

QY 4 EFIFIGGVSSIGKGIASLSLAIEDRGKLYTITLDPYINVDGTMSPQHGGEVFE 63  
 DB 11 YIFITGGVSSIGKGIASLSVLESRLKLYLMDPYINVDGTMSPQHGGEVFE 70  
 QY 64 DGAETDLDLGHYERFLKTTMKNNFTTGQVEOVLNRKGDYLGATVOVIPHITDRIK 123  
 DB 71 DGAETDLDLGHYERFLKTTMKNNFTTGQVEOVLNRKGDYLGATVOVIPHITDRIK 130  
 QY 124 RRVYSABEKDVALLEVGSTVGDIESLPLETRQMGVLEGRDALFHLTLTPYIKSAG 183  
 DB 131 KCLINAGKFDLLEIGSTVGDIESLPLETRQMGVLEGRDALFHLTLTPYIKSAG 190  
 QY 184 ELKTPYQSHVLELRTIGIOPDILCRSBOPIPASERRKIALFTVAEKAIVISADADTI 243  
 DB 191 EYKTPYQSHVLELRTIGIOPDILCRSBOPIPASERRKIALFTVAEKAIVISADADTI 250  
 QY 244 YHPLILLEGQIDDLVVDQLRLDVPADLSAEKVVVDGLTHPTDEVSIAIVGKYVDHTDA 303  
 DB 251 YHPLILLEGQIDDLVVDQLRLDVPADLSAEKVVVDGLTHPTDEVSIAIVGKYVDHTDA 310  
 QY 304 YKSLNEALIHAGIHRHKQVQISYIDSETI-BAEGTAKLKNVDAILVGGGGESEVEXIST 363  
 DB 311 YKSLNEALIHAGIHRHKQVQISYIDSETI-BAEGTAKLKNVDAILVGGGGESEVEXIST 370  
 QY 364 VFAPEENKIPYIGICLQMSAVIEFARNVGLAGASTEPLEKSPHPYIGLITENWDEAG 423

DB 371 AKYSREKIPYIGICLQMSAVIEFARNVGLAGASTEPLEKSPHPYIGLITENWDEAG 430  
 QY 424 ELVTRDESDLGITWELGAOKCEKLSLAFOLYQCDVTEHH 483  
 DB 431 YLNKSNISNNYSTVRLGNQCHLTKGSLAFRIYNSIILERRRAREYNNIFIKI 487  
 QY 484 GKPSGKSLD-----RLVETIILPEHPWFLAOCFHEPFIETPADGHA 536  
 DB 488 --KPSGSLVSGAAYVDNKLVEIIEISNHPFVSGQFHEPFIETPADGHA 545  
 QY 537 KHKQTQ 541  
 DB 546 DFKK 550

RESULT 5  
 ID 08EM53 PRELIMINARY; PRT; 535 AA.  
 AC 08EM53;  
 DT 01-MAR-2003 (TEMBLrel. 23, Created)  
 DT 01-MAR-2003 (TEMBLrel. 23, last sequence update)  
 DE 01-MAR-2003 (TEMBLrel. 23, last annotation update)  
 GN CTP synthase (UTP-ammonia ligase) (EC 6.3.4.2).  
 OS Oceanobacillus iheyensis.  
 OC Bacteria; Firmicutes; Bacillales; Oceanobacillus.  
 NCBI\_TaxID=162710;  
 RX MEDLINE=2220767; PubMed=12235376;  
 RA Takami H., Takaki Y., Uchiyama I.,  
 RT Genome sequence of Oceanobacillus iheyensis isolated from the Iheya  
 RT Ridge and its unexpected adaptive capabilities to extreme  
 RT environments."  
 RL Nucleic Acids Res. 30:3927-3935(2002).  
 DR EMBL; AP004603; BAC14963.1; -  
 KW Ligase; Complete proteome.  
 SQ SEQUENCE 535 AA; 59550 MW; 2EA980BADCT45EC0 CRC64;

Query Match 57.9%; Score 1614; DB 16; Length 535;  
 Best Local Similarity 57.1%; Pred. No. 8.4e-104;  
 Matches 310; Conservative 86; Mismatches 133; Indels 14; Gaps 4;

QY 1 MKKFIPIFGGVSSIGKGIASLSLAIEDRGKLYTITLDPYINVDGTMSPQHGGEVFE 60  
 DB 1 MKKFIPIFGGVSSIGKGIASLSVLESRLKLYLMDPYINVDGTMSPQHGGEVFE 60  
 QY 61 VTODGAETDLDLGHYERFLKTTMKNNFTTGQVEOVLNRKGDYLGATVOVIPHITD 120  
 DB 61 VTODGAETDLDLGHYERFLKTTMKNNFTTGQVEOVLNRKGDYLGATVOVIPHITD 120  
 QY 121 EIKRQVYSAE--GKDVALLHVGSTVGDIESLPLETRQMGVLEGRDALFHLTLTPY 178  
 DB 121 EIKRQVYSAE--GKDVALLHVGSTVGDIESLPLETRQMGVLEGRDALFHLTLTPY 180  
 QY 179 IKSAGELKTPQSHVLELRTIGIOPDILCRSBOPIPASERRKIALFTVAEKAIVISAI 238  
 DB 181 IKSAGELKTPQSHVLELRTIGIOPDILCRSBOPIPASERRKIALFTVAEKAIVISAI 240  
 QY 239 DADITTYIPILLRQGGIDDLVVDQLRLDVPADLSAEKVVVDGLTHPTDEVSIAIVGKY 298  
 DB 241 DADITTYIPILLRQGGIDDLVVDQLRLDVPADLSAEKVVVDGLTHPTDEVSIAIVGKY 300  
 QY 299 DHTDAYSLNEALIHAGIHRHKQVQISYIDSETI-BAEGTAKLKNVDAILVGGGGESEV 357  
 DB 301 ELDPAYLSVASESLHAGYVDVDTVKIHMINSKILSEBQIKELSKYGVAVPGFGDRI 360  
 QY 358 EKGSTVRAPEENKIPYIGICLQMSAVIEFARNVGLAGASTEPLEKSPHPYIGLITE 417  
 DB 361 EKGSTVRAPEENKIPYIGICLQMSAVIEFARNVGLAGASTEPLEKSPHPYIGLITE 420  
 QY 418 WMDAGELVTRDESDLGITWELGAOKCEKLSLAFOLYQCDVTEHH 476

Db 421 QND-----IEDLGLTARIGAPCLVDTAKKAYDAGDVLIEHRHRYENNNY 470  
 Qy 477 LKQLEAAGKFSKSLDGLRLVETIELEPHWPLACQHPDEFTSTPRNGALFSGFVEAA 536  
 Db 471 RQMAKGEVFSQTSFDGRLVETIEVQDHPWVACQHPDEFTSRPRASQLEFGTIGASV 530  
 Qy 537 KKH 539  
 Db 531 DFR 533

## RESULT 6

08G0G1 PRELIMINARY; PRT; 542 AA.  
 ID 08G0G1  
 AC 08G0G1  
 DT 01-MAR-2003 (Tremblrel. 23, Created)  
 DT 01-MAR-2003 (Tremblrel. 23, Last sequence update)  
 DT 01-MAR-2003 (Tremblrel. 23, Last annotation update)  
 DE CTP synthase.  
 GN PYRG OR BR1134.  
 OS Brucella suis.  
 OC Bacteria; Proteobacteria; Alphaproteobacteria; Rhizobiales;  
 OC Brucellaceae; Brucella.  
 OX NCBI\_TaxID=29461;  
 RN (1)  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=1330 / Biovar 1;  
 RX MEDLINE=22247741; PubMed=12271122;  
 RA Paulsen I.T., Seshadri R., Nelson K.E., Eisen J.A., Heidelberg J.F.,  
 RA Read T.D., Dodson R.J., Umayam L., Brinkac L.M., Beanan M.J.,  
 RA Daugherty S.C., Deboy R.T., Durkin A.S., Kolonay J.F., Madupu R.,  
 RA Nelson W.C., Ayodeji B., Krahl M., Shetty J., Malek J., Van Aken S.E.,  
 RA Riedmiller S., Tettelin H., Gill S.R., White O., Salzberg S.L.,  
 RA Hoover D.L., Lindler L.E., Halling S.M., Boyle S.M., Fraser C.M.:  
 RT "The Brucella suis genome reveals fundamental similarities between  
 RT animal and plant pathogens and symbionts."  
 RL Proc. Natl. Acad. Sci. U.S.A. 99:13148-13153(2002).  
 DR EMBL; A014414; AAN30054.1; -.  
 DR TIGR; BR1134; -.  
 KM Complete proteome.  
 SQ SEQUENCE 542 AA; 60058 MW; B88F10406370B7B0 CRC64;

Query Match 56.9%; Score 1585; DB 16; Length 542;  
 Best Local Similarity 58.1%; Pred. No. 8.9e-102;  
 Matches 312; Conservative 78; Mismatches 143; Indels 4; Gaps 4;

Qy 1 MTKFIFITGGVSSSLGKGIASLSAIIEDRGKVTITLDPYINVDPGTMSPOHGEVF 60  
 Db 1 MARYVPIITGVSSSLGKGIASLSAIIEDRGKVTITLDPYINVDPGTMSPOHGEVF 60  
 Qy 61 VTDDGAEFTDLDLGHYERFLKTTMTKNNFTTGVVEQVLENERKGDYLGATVQVPHITD 120  
 Db 61 VTDDGAEFTDLDLGHYERFLKTTMTKNNFTTGVVEQVLENERKGDYLGATVQVPHITD 120  
 Qy 121 EIKRRLVESAS--GKVALIEVGGTGVGDIESLPLETTRQMGVLEGRALFIHLTLVY 180  
 Db 121 EIKRRLVESAS--GKVALIEVGGTGVGDIESLPLETTRQMGVLEGRALFIHLTLVY 180  
 Qy 121 EIKRRLVESAS--GKVALIEVGGTGVGDIESLPLETTRQMGVLEGRALFIHLTLVY 180  
 Db 121 EIKRRLVESAS--GKVALIEVGGTGVGDIESLPLETTRQMGVLEGRALFIHLTLVY 180  
 Qy 181 SAGELTKRGTQHSVKELRTIGIOPDILICRSEOPIPASERRKIALFTTNAEKAVISA 240  
 Db 181 SAGELTKRGTQHSVKELRTIGIOPDILICRSEOPIPASERRKIALFTTNAEKAVISA 240  
 Qy 181 SAGELTKRGTQHSVKELRTIGIOPDILICRSEOPIPASERRKIALFTTNAEKAVISA 240  
 Db 181 SAGELTKRGTQHSVKELRTIGIOPDILICRSEOPIPASERRKIALFTTNAEKAVISA 240  
 Qy 241 DITIRIPLLRREGGLDVLVDQRLD-VPAAIDSAEKVVDGITHETDVSIAIVGKYD 299  
 Db 241 DITIRIPLLRREGGLDVLVDQRLD-VPAAIDSAEKVVDGITHETDVSIAIVGKYD 299  
 Qy 241 DITIRIPLLRREGGLDVLVDQRLD-VPAAIDSAEKVVDGITHETDVSIAIVGKYD 299  
 Db 241 DITIRIPLLRREGGLDVLVDQRLD-VPAAIDSAEKVVDGITHETDVSIAIVGKYD 299  
 Qy 300 HTDAVSLNEALIHAGIHTRHKVOISYIDSEITIEAGTAK-LKNVDAILVPGSGERGVE 358  
 Db 300 HTDAVSLNEALIHAGIHTRHKVOISYIDSEITIEAGTAK-LKNVDAILVPGSGERGVE 358  
 Qy 301 LKDAVSLNEALIHAGIHTRHKVOISYIDSEITIEAGTAK-LKNVDAILVPGSGERGVE 360  
 Db 301 LKDAVSLNEALIHAGIHTRHKVOISYIDSEITIEAGTAK-LKNVDAILVPGSGERGVE 360  
 Qy 359 GKSTYRFARENKIPYLGICLQMSAVIEPARNVGLGASHTEPLPSKPHVIGLITW 418  
 Db 359 GKSTYRFARENKIPYLGICLQMSAVIEPARNVGLGASHTEPLPSKPHVIGLITW 418

Db 361 GKILAAKFAREKRPYFGICFMQACIHAARNLVGIEDASSSERGP-TREPVYGLITW 418  
 Qy 419 MDEAGELVTRDESDSLGSTMELGAOKCLKADSLAFOLYQKOVITERRRHRYEENNYL 478  
 Db 420 L-KGNMELKRAAAGDLGSTMELGAVEALFKEDSKIAQIYGSTDIHRRHRYEENNYL 478  
 Qy 479 QLEAAGKFSKSLDGLRLVETIELEPHWPLACQHPDEFTSTPRNGALFSGFVEAA 535  
 Db 479 RQMAKGEVFSQTSFDGRLVETIEVQDHPWVACQHPDEFTSRPRASQLEFGTIGASV 530

## RESULT 7

08CN12 PRELIMINARY; PRT; 535 AA.  
 ID 08CN12  
 AC 08CN12  
 DT 01-MAR-2003 (Tremblrel. 23, Created)  
 DT 01-MAR-2003 (Tremblrel. 23, Last sequence update)  
 DT 01-MAR-2003 (Tremblrel. 23, Last annotation update)  
 DE CTP synthase.  
 GN SE1725.  
 OS Staphylococcus epidermidis.  
 OC Bacteria; Firmicutes; Bacillales; Staphylococcus.  
 OX NCBI\_TaxID=1282;  
 RN (1)  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=ATCC 12228;  
 RA Zhang Y., Ren S., Li H., Fu G., Lu L., Lu G., Jia J., Tu Y., Qin Z.,  
 RA Chen Z., Wen Y.:  
 RL Submitted (Nov-2002) to the EMBL/Genbank/DBJ databases.  
 DR EMBL; AB016749; AA005324.1; -.  
 KM Complete proteome.  
 SQ SEQUENCE 535 AA; 60211 MW; 629A03087D4AB238 CRC64;

Query Match 56.5%; Score 1575; DB 16; Length 535;  
 Best Local Similarity 55.6%; Pred. No. 4.3e-101;  
 Matches 302; Conservative 95; Mismatches 132; Indels 14; Gaps 4;

Qy 1 MTKFIFITGGVSSSLGKGIASLSAIIEDRGKVTITLDPYINVDPGTMSPOHGEVF 60  
 Db 1 MTKFIFITGGVSSSLGKGIASLSAIIEDRGKVTITLDPYINVDPGTMSPOHGEVF 60  
 Qy 61 VTDDGAEFTDLDLGHYERFLKTTMTKNNFTTGVVEQVLENERKGDYLGATVQVPHITD 120  
 Db 61 VTDDGAEFTDLDLGHYERFLKTTMTKNNFTTGVVEQVLENERKGDYLGATVQVPHITD 120  
 Qy 121 EIKRRLVESAS--GKVALIEVGGTGVGDIESLPLETTRQMGVLEGRALFIHLTLVY 178  
 Db 121 EIKRRLVESAS--GKVALIEVGGTGVGDIESLPLETTRQMGVLEGRALFIHLTLVY 178  
 Qy 121 EIKRRLVESAS--GKVALIEVGGTGVGDIESLPLETTRQMGVLEGRALFIHLTLVY 178  
 Db 121 EIKRRLVESAS--GKVALIEVGGTGVGDIESLPLETTRQMGVLEGRALFIHLTLVY 178  
 Qy 179 IKSAGELTKRGTQHSVKELRTIGIOPDILICRSEOPIPASERRKIALFTTNAEKAVISA 238  
 Db 179 IKSAGELTKRGTQHSVKELRTIGIOPDILICRSEOPIPASERRKIALFTTNAEKAVISA 238  
 Qy 181 IKSAGELTKRGTQHSVKELRTIGIOPDILICRSEOPIPASERRKIALFTTNAEKAVISA 240  
 Db 181 IKSAGELTKRGTQHSVKELRTIGIOPDILICRSEOPIPASERRKIALFTTNAEKAVISA 240  
 Qy 229 DADTIRIPLLRREGGLDVLVDQRLDVP-AADLSAEKVVDDGITHETDVSIAIVGKY 297  
 Db 229 DADTIRIPLLRREGGLDVLVDQRLDVP-AADLSAEKVVDDGITHETDVSIAIVGKY 297  
 Qy 241 DADTIRIPLLRREGGLDVLVDQRLDVP-AADLSAEKVVDDGITHETDVSIAIVGKY 300  
 Db 241 DADTIRIPLLRREGGLDVLVDQRLDVP-AADLSAEKVVDDGITHETDVSIAIVGKY 300  
 Qy 298 VDRDAYSINLEALIHAGIHTRHKVOISYIDSEITIEAGT-AKLKNVDAILVPGSGERG 356  
 Db 298 VDRDAYSINLEALIHAGIHTRHKVOISYIDSEITIEAGT-AKLKNVDAILVPGSGERG 356  
 Qy 301 VDRDAYSINLEALIHAGIHTRHKVOISYIDSEITIEAGT-AKLKNVDAILVPGSGERG 360  
 Db 301 VDRDAYSINLEALIHAGIHTRHKVOISYIDSEITIEAGT-AKLKNVDAILVPGSGERG 360  
 Qy 357 VEGKISTYRFARENKIPYLGICLQMSAVIEPARNVGLGASHTEPLPSKPHVIGLIT 416  
 Db 357 VEGKISTYRFARENKIPYLGICLQMSAVIEPARNVGLGASHTEPLPSKPHVIGLIT 416  
 Qy 361 SEGKIAATRYARENNIPYLGICLQMSAVIEPARNVGLGASHTEPLPSKPHVIGLIT 420  
 Db 361 SEGKIAATRYARENNIPYLGICLQMSAVIEPARNVGLGASHTEPLPSKPHVIGLIT 420  
 Qy 417 EMDGAGELVTRDESDSLGSTMELGAOKCLKADSLAFOLYQKOVITERRRHRYEENNY 476  
 Db 417 EMDGAGELVTRDESDSLGSTMELGAOKCLKADSLAFOLYQKOVITERRRHRYEENNY 476  
 Qy 421 EQKQ-----IEDLGLTARIGAPCLVDTAKKAYDAGDVLIEHRHRYEENNY 470  
 Db 421 EQKQ-----IEDLGLTARIGAPCLVDTAKKAYDAGDVLIEHRHRYEENNY 470  
 Qy 477 LKQLEAAGKFSKSLDGLRLVETIELEPHWPLACQHPDEFTSTPRNGALFSGFVEAA 536  
 Db 477 LKQLEAAGKFSKSLDGLRLVETIELEPHWPLACQHPDEFTSTPRNGALFSGFVEAA 536  
 Qy 479 QLEAAGKFSKSLDGLRLVETIELEPHWPLACQHPDEFTSTPRNGALFSGFVEAA 535  
 Db 479 QLEAAGKFSKSLDGLRLVETIELEPHWPLACQHPDEFTSTPRNGALFSGFVEAA 535

QY 537 RHK 539  
Db 531 NYQ 533

## RESULT 8

QBDQYO PRELIMINARY; PRT; 535 AA.

AC QBDQYO;  
DT 01-MAR-2003 (TREMblrel. 23, Created)  
DT 01-MAR-2003 (TREMblrel. 23, Last sequence update)  
DE CTP synthetase (EC 6.3.4.2).  
GN PYRG OR SPR0438.  
OS Streptococcus pneumoniae (strain ATCC BAA-255 / R6).  
OC Bacteria; Firmicutes; Lactobacillales; Streptococcaceae;  
OC Streptococcus.  
OX NCBI\_TaxID=171101;  
RN [1]  
RP SEQUENCE FROM N.A.  
RX MEDLINE=21429245; PubMed=11544234;  
RA Hoskins J., Albom W.E. Jr., Arnold J., Blaszcak L.C., BURGESS S.,  
RA Dehoff B.S., Estrem S.T., Fritz L., Fu D.-J., Fuller W., Geringer C.,  
RA Gilmour R., Glass J.S., Kioja H., Kraft A.R., Lagace R.E.,  
RA LeBlanc D.J., Lee L.N., Leikowitz E.J., Lu J., Matsushima P.,  
RA McArthur S.M., McHenry M., Mcleaster K., Mundy C.W., Nicas T.I.,  
RA Norris F.H., O'Gara M., Peery R.B., Robertson G.T., Rokey P.,  
RA Sun P.-M., Winkler M.E., Yang Y., Young-Bellido M., Zhao G.,  
RA Zook C.A., Baltz R.H., Jaskunas S.R., Rostek P.R. Jr., Skatard P.L.,  
RA Glass J.I.,  
RT "Genome of the bacterium Streptococcus pneumoniae strain R6."  
RL J. Bacteriol. 183:5709-5717(2001).  
DR EMBL; AB008423; AAK9242.1; -  
KM Ljagee; Complete proteome.  
SQ SEQUENCE 535 AA; 59256 MW; 7BC6C267AF74C12 CRC64;

Query Match 56.0%; Score 1561; DB 16; Length 535;  
Best Local Similarity 54.3%; Pred. No. 4.1e-100;  
Matches 294; Conservative 108; Mismatches 125; Indels 14; Gaps 4;

2 TKEFTFTGCVSSIGKGIASSLAILEDRLKATTTKIDPYINVDPGTMSPPQGEVY 61  
Db 3 TKITFTGCVSSIGKGIASSLAILEDRLKATTTKIDPYINVDPGTMSPPQGEVY 62  
QY 62 TDEGAEITDLDLGHYERFLKTTKNNFTTGYEYVLRNRRKDYLGATVQYIPIHTD 121  
Db 63 TDDGAEITDLDLGHYERFLKTTKNNFTTGYEYVLRNRRKDYLGATVQYIPIHTD 122  
QY 122 IKRRVYSSA--EGKDVALLIEVGTVGDIIESLPLETTIRKMGVSELGRDALFIHTLVPI 179  
Db 123 LKEKIRRAALTDSVATTEVGTVGDIIESLPLETTIRKMGVSELGRDALFIHTLVPI 182  
QY 180 KSAGEELTKTPQHSVSKELRTIGIOPDILICRSEOPIPASERRKIALFTNVAEKAVISAID 239  
Db 183 KAAGEEMTKTPQHSVSKELRTIGIOPDILICRSEOPIPASERRKIALFTNVAEKAVISAID 242  
QY 240 ADITTRIPILLRBOGLDLDLVDDQLRDVPADLSANEKVYDGLTHPTDVSIAIVGKYD 293  
Db 243 VEHLYQIPNLQKQGMDOIVCDHLKLDAPADMTESAMVDKWNKKQKSLVKEYE 302  
QY 300 HTDAYSINLEALIHAGIHERHKKVQISYIDSETIEAGTAK-LKNVDAILVPGSGGEGVY 358  
Db 303 LQDAIVSYVVALHSGVNVNVEKIMVNNNDVTAEVVALISDADGIIYVPGFGRGTE 362  
QY 359 GKISTVPAFAENKIPYIGICLQMSAVTERANVVGLEGASTFPLKSPHPYIGLITE 418  
Db 363 GKIQARVAENENVPMLGVCLQWQCTICIFARHVLGLEGANSALPEKYPPIIDMRQ 422  
QY 419 MDAEGLVTRDEDSDLGITRLGAQKRLKADSLAFOLY-QKDVITERRRHRYEFNNQYL 477  
Db 423 ID-----IDMGITRLGLIYPSKLRGSKAAAAAHNQEYVQRHRRHRYEFNNAR 472  
QY 478 KOLEAAGMKFSKSLDGLVLEIIEPPEWFLACQHPHFTSTPRNGHALPSGVEAAAK 537

Db 473 EGFPAAGPFFSGVSPDNRLVLEIIEPPEWFLACQHPHFTSTPRNGHALPSGVEAAAK 532  
QY 538 H 538  
Db 533 N 533

## RESULT 9

QBDKT7 PRELIMINARY; PRT; 543 AA.

AC QBDKT7;  
DT 01-MAR-2003 (TREMblrel. 23, Created)  
DT 01-MAR-2003 (TREMblrel. 23, Last sequence update)  
DE CTP synthetase.  
GN PYRG OR TL0768.  
OS Synecchococcus elongatus (Thermosynechococcus elongatus).  
OC Bacteria; Cyanobacteria; Chroococcales; Synecchococcus.  
OX NCBI\_TaxID=32046;  
RN [1]  
RP SEQUENCE FROM N.A.  
RX STRAIN=BP-1;  
RX MEDLINE=2225144; PubMed=12240834;  
RA Nakamura Y., Kaneko T., Sato S., Ikeuchi M., Katoh H., Saeamoto S.,  
RA Watanabe A., Iriiguchi M., Kawashima K., Kimura T., Kishida Y.,  
RA Kiyokawa C., Kohara M., Matsumoto M., Matsuno A., Nakazaki N.,  
RA Shimo S., Sugimoto M., Takeuchi C., Yamada M., Tabata S.;  
RT "Complete genome structure of the thermophilic cyanobacterium  
Thermosynechococcus elongatus BP-1."  
RL DNA Res. 9:123-130(2002).  
DR EMBL; AP005371; BAC08319.1; -  
KM Complete proteome.  
SQ SEQUENCE 543 AA; 60484 MW; 0884DA82309EFEE40 CRC64;

Query Match 55.8%; Score 1555.5; DB 16; Length 543;  
Best Local Similarity 57.0%; Pred. No. 1e-99;  
Matches 310; Conservative 80; Mismatches 141; Indels 13; Gaps 3;

1 MTKFTFTGCVSSIGKGIASSLAILEDRLKATTTKIDPYINVDPGTMSPPQGEVY 60  
Db 1 MKTYFTGCVSSIGKGIASSLAILEDRLKATTTKIDPYINVDPGTMSPPQGEVY 60  
QY 61 VTDEGAEITDLDLGHYERFLKTTKNNFTTGYEYVLRNRRKDYLGATVQYIPIHTD 120  
Db 61 VTDDGAEITDLDLGHYERFLKTTKNNFTTGYEYVLRNRRKDYLGATVQYIPIHTD 120  
QY 121 IKRRVYSSABGK--DVALLIEVGTVGDIIESLPLETTIRKMGVSELGRDALFIHTLVPI 178  
Db 121 EIKERILRAKKNPDVYIIEIGTVGDIIESLPLETTIRKMGVSELGRDALFIHTLVPI 180  
QY 179 KSAGEELTKTPQHSVSKELRTIGIOPDILICRSEOPIPASERRKIALFTNVAEKAVISAID 238  
Db 181 ISSAGEEMTKTPQHSVSKELRTIGIOPDILICRSEOPIPASERRKIALFTNVAEKAVISAID 240  
QY 239 DADITTRIPILLRBOGLDLDLVDDQLRDVPADLSANEKVYDGLTHPTDVSIAIVGKYD 298  
Db 241 DKSITVYEPFLRERGLAVQVNLNLLEOROPDLSQWQCTICIFARHVLGSHRPLEVAAIKV 300  
QY 299 DHTDAYSINLEALIHAGIHERHKKVQISYIDSETIEAGTAK-ANKVNDAILVPGSGGEGVY 357  
Db 301 RLSDAIVSYVVALHSGVNVNVEKIMVNNNDVTAEVVALISDADGIIYVPGFGRGTE 360  
QY 359 GKISTVPAFAENKIPYIGICLQMSAVTERANVVGLEGASTFPLKSPHPYIGLITE 417  
Db 361 EKIKAIVYARQGIPIPIGLCLQWQCAVIEWRHVLGLEGANSALPEKYPPIIDMRQ 420  
QY 418 MDAEGLVTRDEDSDLGITRLGAQKRLKADSLAFOLY-QKDVITERRRHRYEFNNQYL 477  
Db 421 QQDIY-----DAGITRLGLIYACRLAHSLAEKLYGSTVIVYERHRRHRYEFNNAR 470  
QY 478 KOLEAAGMKFSKSLDGLVLEIIEPPEWFLACQHPHFTSTPRNGHALPSGVEAAAK 537

DB 471 NLEFETGYQITGTS PDGRLVEIIEYPAHPEFIAVGFHREBRSPNAPHPLFYGLLAAAK 530  
 QY 538 HKTQ 541  
 DB 531 NSNR 534

## RESULT 10

Q8E7P8 PRELIMINARY; PRT: 534 AA.  
 ID Q8E7P8; AC Q8E7P8;  
 DT 01-MAR-2003 (TrEMBLrel. 23, Created)  
 DT 01-MAR-2003 (TrEMBLrel. 23, Last sequence update)  
 DT 01-MAR-2003 (TrEMBLrel. 23, Last annotation update)  
 DE Hypothetical protein.  
 GN GRS0106.  
 OS Streptococcus agalactiae (serotype III).  
 OC Bacteria; Firmicutes; Lactobacillales; Streptococcaceae;  
 OC Streptococcus.  
 OX NCBI\_TaxID=216495;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=MEM316 / Serotype III;  
 RX MEDLINE=22242508; PubMed=12354221;  
 RA Glaeser P., Rustiok C., Buchrieser C., Chevalier F., Frangoul L.,  
 RA Msadek T., Zouine M., Couve E., Talloni L., Poyart C., Titeu-Cuot P.,  
 RA Kunst F.;  
 RT "Genome sequence of Streptococcus agalactiae, a pathogen causing  
 RT invasive neonatal disease."  
 RL Mol. Microbiol. 45:1499-1513(2002).  
 DR EMBL: AL766843; CAD45751.1; -;  
 DR Sagalistic; gprs0106; -;  
 KW Hypothetical protein; Complete proteome.  
 SQ SEQUENCE 534 AA; 59294 MW; 0CE1CC8511EDB5 CRC64;

Query Match 55.0%; Score 1533; DB 16; Length 534;  
 Best Local Similarity 53.1%; Pred. No. 3.6e-98;  
 Matches 288; Conservative 108; Mismatches 132; Indels 14; Gaps 4;

QY 1 MTKYIFITGGVSSIGKGIASSLAAILDRGLKVTITKLPYINVDPTGMSPGHGEV 60  
 DB 1 MTKYIFITGGVSSIGKGIASSLAAILDRGLKVTITKLPYINVDPTGMSPGHGEV 60  
 QY 61 VTDGAEITDLDGHERFLKTKMTKKNFTTGOVBOYLREKGDYLGATVOYPIHTD 120  
 DB 61 VTDGAEITDLDGHERFLKTKMTKKNFTTGOVBOYLREKGDYLGATVOYPIHTD 120  
 QY 121 EIKRVEESA--EGKVALIEVGTVGDIESTLPLETIRQMGVELGRDALFIHLLVY 178  
 DB 121 ALKEKIKRAATTDSVITTEVGTVGDIESTLPLETIRQMGVELGRDALFIHLLVY 180  
 QY 179 IKSAGELKTKTQSHSVKELRTIGIOPDILICRSEQPIPAERKIALFTNVAEKAVISAI 238  
 DB 181 LKAAGMKTKPTQSHSVKELRTIGIOPDILICRSEQPIPAERKIALFTNVAEKAVISAI 240  
 QY 239 DADTIRIFILLREOGDLDLVVDQLRDVPAADLSAEKRVDPGLTPTDEVSIAIVGKY 298  
 DB 241 DVDHITQIPLNMQQONNDQIYCDHLKLETPADMTKMSAVDKRMNLEKTKYKIALVGY 300  
 QY 299 DHTDAVYSLNHALTHAGIHTRHAKVQIYISSETIEAGTAKL-KNVDAILVPGGFERGV 357  
 DB 301 ELPDALISVEALIKSHGYNDVAIDLKVNAAVEEDNIKELVGDAGIIVPGGFGGRGS 360  
 QY 358 EKGISTVREARENKIPYIGICIGQSAVIEFARNVVGEGHSTPELPKSPHPYIGLITE 417  
 DB 361 EKGISTVREARENKIPYIGICIGQSAVIEFARNVVGEGHSTPELPKSPHPYIGLITE 420  
 QY 418 WNDDEGELVTRDSDSLGGTMRIGAKCKLADSLAPQY-OKDVIYTERHRRHRYEPNNQY 476  
 DB 421 QID-----IEDMGSTLRIGLYPCILKAGSSAAAYNNQVAVQVRRHRRHRYEPNNQY 470  
 QY 477 LKQLBAAGKFFSGKSLDGLVETIELPEHPMFLAOCFFHEFTSTRNGHALPSGVEAAA 536  
 DB 477 LKQLBAAGKFFSGKSLDGLVETIELPEHPMFLAOCFFHEFTSTRNGHALPSGVEAAA 536

DB 471 RQPEAAGVFGVSGPDRNLMVEVELPEKKPFVAQAQYHPELQSRPNHAEELTAVTAAY 530  
 QY 537 KH 538  
 DB 531 EN 532

## RESULT 11

Q8E290 PRELIMINARY; PRT: 534 AA.  
 ID Q8E290; AC Q8E290;  
 DT 01-MAR-2003 (TrEMBLrel. 23, Created)  
 DT 01-MAR-2003 (TrEMBLrel. 23, Last sequence update)  
 DT 01-MAR-2003 (TrEMBLrel. 23, Last annotation update)  
 DE CTP synthase  
 GN PYR3 OR SAC0107.  
 OS Streptococcus agalactiae (serotype V).  
 OC Bacteria; Firmicutes; Lactobacillales; Streptococcaceae;  
 OC Streptococcus.  
 OX NCBI\_TaxID=216466;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=2603 V/R / Serotype V;  
 RX MEDLINE=22222988; PubMed=12200547;  
 RA Tettelein H., Masiyann V., Cieslewicz M.J., Eisen J.A., Peterson S.,  
 RA Wesels M.R., Paulsen I.T., Nelson K.E., Margalit I., Read T.D.,  
 RA Madoff L.C., Wolf A.M., Beaman M.J., Brinkac L.M., Daugherty S.C.,  
 RA DeBoy R.T., Durkin A.S., Kolman J.F., Madupu R., Lewis M.R.,  
 RA Radune D., Fedorova N.B., Scanlan D., Khouri H., Mulligan S.,  
 RA Caray H.A., Cline R.T., Van Aken S.E., Gill J., Scarselli M., Mora M.,  
 RA Jacobini E.T., Bretton C., Gall G., Mariani M., Vegni F., Malone D.,  
 RA Rinaldo D., Rappelli R., Telford J.L., Kasper D.L., Grandi G.,  
 RA Frazer C.M.;  
 RT "Complete genome sequence and comparative genomic analysis of an  
 RT emerging human pathogen, serotype V Streptococcus agalactiae."  
 RL Proc. Natl. Acad. Sci. U.S.A. 99:12391-12396(2002).  
 DR EMBL: AE014196; AAM99015.1; -;  
 DR TIGR: SAG0107; -;  
 KW Complete proteome.  
 SQ SEQUENCE 534 AA; 59296 MW; 542AD55C132CC870 CRC64;

Query Match 54.9%; Score 1532; DB 16; Length 534;  
 Best Local Similarity 53.0%; Pred. No. 4.2e-98;  
 Matches 287; Conservative 109; Mismatches 132; Indels 14; Gaps 4;

QY 1 MTKYIFITGGVSSIGKGIASSLAAILDRGLKVTITKLPYINVDPTGMSPGHGEV 60  
 DB 1 MTKYIFITGGVSSIGKGIASSLAAILDRGLKVTITKLPYINVDPTGMSPGHGEV 60  
 QY 61 VTDGAEITDLDGHERFLKTKMTKKNFTTGOVBOYLREKGDYLGATVOYPIHTD 120  
 DB 61 VTDGAEITDLDGHERFLKTKMTKKNFTTGOVBOYLREKGDYLGATVOYPIHTD 120  
 QY 121 EIKRVEESA--EGKVALIEVGTVGDIESTLPLETIRQMGVELGRDALFIHLLVY 178  
 DB 121 ALKEKIKRAATTDSVITTEVGTVGDIESTLPLETIRQMGVELGRDALFIHLLVY 180  
 QY 179 IKSAGELKTKTQSHSVKELRTIGIOPDILICRSEQPIPAERKIALFTNVAEKAVISAI 238  
 DB 181 LKAAGMKTKPTQSHSVKELRTIGIOPDILICRSEQPIPAERKIALFTNVAEKAVISAI 240  
 QY 239 DADTIRIFILLREOGDLDLVVDQLRDVPAADLSAEKRVDPGLTPTDEVSIAIVGKY 298  
 DB 241 DVDHITQIPLNMQQONNDQIYCDHLKLETPADMTKMSAVDKRMNLEKTKYKIALVGY 300  
 QY 299 DHTDAVYSLNHALTHAGIHTRHAKVQIYISSETIEAGTAKL-KNVDAILVPGGFERGV 357  
 DB 301 ELPDALISVEALIKSHGYNDVAIDLKVNAAVEEDNIKELVGDAGIIVPGGFGGRGS 360  
 QY 358 EKGISTVREARENKIPYIGICIGQSAVIEFARNVVGEGHSTPELPKSPHPYIGLITE 417  
 DB 361 EKGISTVREARENKIPYIGICIGQSAVIEFARNVVGEGHSTPELPKSPHPYIGLITE 420

QY 418 WMDAGELVTRDSDSLGTMRLGAOKCRKADSLAFOLY-QKDYTERHRHREFFNNQY 476  
DB 421 QID-----IDMGGLRLGLYPCCKKSGSRAPAAANNVQRRHRHREFFNNQY 470  
QY 477 LKQLEAAGKFGSKSLDGLRVEILPEHFWPLACOFHEFTSTPRNGHALFSGFEVAAA 536  
DB 471 REQPEAAGFVSGVSPDRMLMEVVELPEKCFVAAQVHPELGRFVHABEVLTAFAVTA 530  
QY 537 KH 538  
DB 531 EN 532  
RESULT 12  
Q8F3J3  
ID Q8F3J3 PRELIMINARY; PRT; 542 AA.  
AC Q8F3J3;  
DT 01-MAR-2003 (Tremblrel. 23, Created)  
DT 01-MAR-2003 (Tremblrel. 23, last sequence update)  
DT 01-MAR-2003 (Tremblrel. 23, last annotation update)  
DE CTP synthase (EC 6.3.4.2).  
GN PYRG OR LA2409.  
OS Leptospira interrogans.  
OC Bacteria; Spirochaetes; Spirochaetales; Leptospiraceae; Leptospira.  
NCBI\_TaxID=173;  
RN NCBI\_TaxID=173;  
RP SEQUENCE FROM N.A.  
RA STRAIN=56601 / Serogroup Icterohaemorrhagiae / Serovar 1a1;  
RL Submitted (MAR-2002) to the EMBL/GenBank/DBJ databases.  
KW Ligase; Complete proteome.  
SQ SEQUENCE 542 AA; 61024 MW; 80BF9DBF864AF8A CRC64;  
Query Match 54.9%; Score 1532; DB 16; Length 542;  
Best Local Similarity 53.4%; Pred. No. 4.3e-98;  
Matches 221; Conservative 99; Mismatches 141; Indels 14; Gaps 4;  
QY 2 TXRIFITGVVSSLGKGIASSLAILEDRLKVTITKLDPIYINVDPTGMSPFQHGVEV 61  
DB 8 TXRIFITGVVSSLGKGIASSLAILEDRLKVTITKLDPIYINVDPTGMSPFQHGVEV 67  
QY 62 TEDGATDLDLGHYERFLKTTMKNNFTTGQYEVOLRNERRKDYLGATVOYIPIHIDE 121  
DB 68 TADGATDLDLGHYERFLKTTMKNNFTTGQYEVOLRNERRKDYLGATVOYIPIHIDE 127  
QY 122 IKRRVESA--EGKDVALLIEVGTVGDIISLPLETRIRKMGVGLGRDALFIHLTLVPI 179  
DB 128 IKRRKTIYARENDPDIIVEIGTVGDIISLPLETRIRKMGVGLGRDALFIHLTLVPI 187  
QY 180 KSAGELTKPTQHSVKELRTIGIOPDILICRSEOPIPASERRKIALFTNVAEKAVISAI 239  
DB 188 TAAAGAKTKPTQHSVKELRTIGIOPDILICRSEOPIPASERRKIALFTNVAEKAVISAI 247  
QY 240 ADT-ITRILLRBOGLDLDVDDRLDVPADLSMERVVDGLTHPTDVSIAIVGRV 298  
DB 248 ISTSITIEIKTKYEEKLDDEVLTAKMEELRSNFSFSDMKVKGILLTKQVLAIVAGKI 307  
QY 299 DHTDAVKSINELIHAGIHTRRKVOISYIDSETIIEAGTAK-LKNVDAILVGGGGERGV 357  
DB 308 SIQDAVRSIYESSHSGIAHDTKEFIKYDPENLNDQSYVELLKVHGLIVGGGGERGI 367  
QY 358 EKKISTVPAERENKIPYLGICLGMQSAVIEPARNVVGLGASHSTFPLKSPHPYIGLITE 417  
DB 368 EKKITAIQVARNNGIPFGICLGMQSAVIEPARNVVGLGASHSTFPLKSPHPYIGLITE 427  
QY 418 WMDAGELVTRDSDSLGTMRLGAOKCRKADSLAFOLY-QKDYTERHRHREFFNNQY 476  
DB 428 -----QNDIEMSGTMRLLGSPCKVAKENTISYSYKSIILHERRHRHREFFNNQY 477  
QY 478 KQLEAAGKFGSKSLDGLRVEILPEHFWPLACOFHEFTSTPRNGHALFSGFEVAAA 536  
DB 478 KQYEEAGFVSGVSPDRMLMEVVELPEKCFVAAQVHPELGRFVHABEVLTAFAVTA 530

QY 538 HKTQG 542  
DB 538 YSKKG 542  
RESULT 13  
Q8DWG1  
ID Q8DWG1 PRELIMINARY; PRT; 536 AA.  
AC Q8DWG1;  
DT 01-MAR-2003 (Tremblrel. 23, Created)  
DT 01-MAR-2003 (Tremblrel. 23, last sequence update)  
DT 01-MAR-2003 (Tremblrel. 23, last annotation update)  
DE CTP synthase (UTP-ammonia lyase) (EC 6.3.4.2).  
GN PYRG OR SMU.97.  
OS Streptococcus mutans.  
OC Bacteria; Firmicutes; Lactobacillales; Streptococcaceae;  
OC Streptococcus.  
NCBI\_TaxID=1309;  
RN NCBI\_TaxID=1309;  
RP SEQUENCE FROM N.A.  
RA STRAIN=UAI59 / ATCC 700610 / Serotype C;  
RC MEDLINE=22295063; PubMed=12397186;  
RA Ajdic D., McShan W.M., McLaughlin R.E., Savic G., Chang J.,  
RA Carson M.B., Primeaux C., Tian R., Kenton S., Jia H., Lin S., Qian Y.,  
RA Li S., Zhu H., Najjar F., Lai H., White J., Roe B.A., Ferretti J.J.  
RT "Genome sequence of Streptococcus mutans UAI59, a cariogenic dental  
RT pathogen."  
RT Proc. Natl. Acad. Sci. U.S.A. 99:14434-14439 (2002).  
RL EMBL; AE014861; AAN57880.1;  
KW Ligase; Complete proteome.  
SQ SEQUENCE 536 AA; 59491 MW; 1512BADCEBDF0446 CRC64;  
Query Match 54.3%; Score 1514; DB 16; Length 536;  
Best Local Similarity 53.1%; Pred. No. 7.6e-97;  
Matches 288; Conservative 102; Mismatches 138; Indels 14; Gaps 4;  
QY 1 MTXRFITGVVSSLGKGIASSLAILEDRLKVTITKLDPIYINVDPTGMSPFQHGVEV 60  
DB 1 MTXRFITGVVSSLGKGIASSLAILEDRLKVTITKLDPIYINVDPTGMSPFQHGVEV 60  
QY 61 VTDDGATDLDLGHYERFLKTTMKNNFTTGQYEVOLRNERRKDYLGATVOYIPIHIDE 120  
DB 61 VTDDGATDLDLGHYERFLKTTMKNNFTTGQYEVOLRNERRKDYLGATVOYIPIHIDE 120  
QY 121 EIKRRVESA--EGKDVALLIEVGTVGDIISLPLETRIRKMGVGLGRDALFIHLTLVPI 178  
DB 121 ALKTKIRKAAATTTSDVITITVIGTVGDIISLPLETRIRKMGVGLGRDALFIHLTLVPI 180  
QY 179 IKSAGELTKPTQHSVKELRTIGIOPDILICRSEOPIPASERRKIALFTNVAEKAVISAI 238  
DB 181 LKAAGEMTKPTQHSVKELRTIGIOPDILICRSEOPIPASERRKIALFTNVAEKAVISAI 240  
QY 239 DADTIYRILLRBOGLDLDVDDRLDVPADLSMERVVDGLTHPTDVSIAIVGRV 298  
DB 241 DVDDYQIPPLVQKQNDQIVCDHLKDVPAADMTSASAMVDKWNLLKCKTKKIALVGRV 300  
QY 299 DHTDAVKSINELIHAGIHTRRKVOISYIDSETIIEAGTAK-TAKLKNVDAILVGGGGERGV 357  
DB 301 ELPAVILSVLEALHSGVNDPTALDMMINANEVTPPEVTDLDGADGIIIVPGGGERGT 360  
QY 358 EKKISTVPAERENKIPYLGICLGMQSAVIEPARNVVGLGASHSTFPLKSPHPYIGLITE 417  
DB 361 EKKIETARYARENDVPMGLICLGMQSAVIEPARNVVGLGASHSTFPLKSPHPYIGLITE 420  
QY 418 WMDAGELVTRDSDSLGTMRLGAOKCRKADSLAFOLY-QKDYTERHRHREFFNNQY 476  
DB 421 QID-----IDMGGLRLGLYPCCKKSGSKTSAYSNQGVQRRHRHREFFNNQY 470  
QY 477 LKQLEAAGKFGSKSLDGLRVEILPEHFWPLACOFHEFTSTPRNGHALFSGFEVAAA 536  
DB 471 RQPEEAGFVSGVSPDRMLMEVVELPEKCFVAAQVHPELGRFVHABEVLTAFAVTA 530

QY 537 KH 538  
DB 531 EN 532

## RESULT 14

Q865X7 PRELIMINARY; PRT; 553 AA.  
ID 0865X7  
AC 0865X7  
DT 01-MAR-2003 (TREMblrel. 23, Created)  
DT 01-MAR-2003 (TREMblrel. 23, Last sequence update)  
DT 01-MAR-2003 (TREMblrel. 23, Last annotation update)  
DE CTP synthase.  
OS PYRG OR BL0874.  
OC Bifidobacterium longum.  
OC Bacteria: Actinobacteriia: Actinobacteridae; Bifidobacteriales;  
OC Bifidobacteriaceae; Bifidobacterium.  
OX NCBI\_TaxID=216816;  
RN (1)  
RP SEQUENCE FROM N.A.  
RC STRAIN=NC 2705;  
RX MEDLINE=22324977; PubMed=12381787;  
RA Schell M.A., Karmaliantzou M., Snel B., Vlianova D., Berger B.,  
RA Pessi G., Zwaan M.C., Desiere F., Bork P., Delley M.,  
RA Pridmore R.D., Arigon F.;  
RT "The genome sequence of Bifidobacterium longum reflects its adaptation  
RT to the human gastrointestinal tract."  
RL Proc. Natl. Acad. Sci. U.S.A. 99:14422-14427(2002).  
DR EMBL; AB014709; AAN24687.1; -  
KW Complete proteome.  
SQ SEQUENCE 553 AA; 60983 MW; 8C8EA4ABECCF48C0 CRC64;

Query Match 51.4%; Score 1432.5; DB 16; Length 553;  
Best Local Similarity 52.3%; Pred. No. 3.7e-91;  
Matches 285; Conservative 97; Mismatches 150; Indels 13; Gaps 6;

QY 1 MKKFFITGCVSSSLGKGLAASSLAILEDRLKTTITLQDPIYINVDGTSPPFGHGYF 60  
DB 13 VTKH:FYVGVGVSSLGKGLTASSLRSLRSRGIHVLQOQDPIYINVDGTSPPFGHGYF 72  
QY 61 VEDGAEETDLDGHERFLKTKKNNFTTQGYEQLVNRKGDYLGATVQVPHITD 120  
DB 73 VEDGAEETDLDGHERFLDVLISQKANTTQGIQVYLRKRAHAYLQGCQVPHITD 132  
QY 121 EIKRRV-YESAGKQVALIEVGTVGDIESLPPELTIQVGEIGRDLPIHLTVPI 179  
DB 133 EIKSRMRQASDDVYVITEIGTVGDIESQPLEAAREVRDLGPRDCMFHVSIVPEYI 192  
QY 180 KSAEGLKTPQHVSVAZETIGIOPDILICRSEOPIPASERKIALFTVMKAVISAID 239  
DB 193 SAHHEKTYPTQHSVMMRLQGLISPDALVNSDRPLNQS IKDKISLMCDVADAGVNCVD 252  
QY 240 ADTIYRIPLLAREQGLDLYVDQLRDVPAADLSAMEKVVGLTHTPDEVSAIVGKYVD 299  
DB 253 ABSIYDVPKILPEEGDADVYVRELGLPFHDVDMENADLLEVRHHPKHEVNAIVGKYID 312  
QY 300 HTDAYSLSNEALIHAGIHTRHVVOISYIDSEITE-AEG-TALQKAVDALVPGGEGGV 357  
DB 313 LPDAVLSVEALKAQGANMAVAVKVAADACETTEGAALADLVVDGIVIPGGGIGI 372  
QY 358 ESKISTVFARENKIPYLGICLQMSAVIEFARNVVGLEGASTFLPKSPHPVGLITE 417  
DB 373 DCKIGLKARARETKLPALGLCIGLSQWVIEGRHVLGIEDANSSFEEDCANPIYATMBE 432  
QY 418 WMD-EAGELVTEDESDIGCTMELAQCKRLKADSLAQOLYQKQVITERRHRYEFNNQY 476  
DB 433 QMDIYAGK-----GDMGHTMRLGSSYPALLEGSLVLELGTIVTERHRRREVNAY 485  
QY 477 LKQLRAAGKPSGKSLDGLVLEIIELEP--HPMFLACQFHPFTSTPRNGALPSGFA 534  
DB 486 KDLREGGIRISGQSPDGLTFLVELPDQVHPFYATQAHPEFKRPTKPHPLPAGLVKA 545  
QY 535 AAHKH 539

DB 546 ALDHQ 550

## RESULT 15

Q86D26 PRELIMINARY; PRT; 468 AA.  
ID 086D26  
AC 086D26  
DT 01-MAR-2003 (TREMblrel. 23, Created)  
DT 01-MAR-2003 (TREMblrel. 23, Last sequence update)  
DT 01-MAR-2003 (TREMblrel. 23, Last annotation update)  
DE CTP synthase (EC 6.3.4.2) (Fragment).  
OS Helicobacterium mobilis.  
OC Bacteria: Firmicutes; Clostridia; Clostridiales; Helicobacteriaceae;  
OC Helicobacterium.  
OX NCBI\_TaxID=28064;  
RN (1)  
RP SEQUENCE FROM N.A.  
RX MEDLINE=22337798; PubMed=12446909;  
RA Raymond J., Zhaxybayeva O., Gogarten J.P., Gerdes S.Y.,  
RA Blankenship R.E.;  
RT "Whole-genome analysis of photosynthetic prokaryotes."  
RL Science 298:1616-1620(2002).  
RN (2)  
RP SEQUENCE FROM N.A.  
RA Lidjols K.G., Chu L., Ostrovskaya O., Mendybaeva N., Koukharenko V.,  
RA Gerdes S., Kyrides N., Overbeek R.;  
RL Submitted (AUG-2002) to the EMBL/GenBank/DBJ databases.  
DR EMBL; AY142825; AAN87429.1; -  
KW Ligase.  
KM NON TER  
SQ SEQUENCE 468 AA; 52317 MW; 51E548C87D98AAS CRC64;

Query Match 50.8%; Score 1416.5; DB 2; Length 468;  
Best Local Similarity 57.7%; Pred. No. 3.7e-90;  
Matches 276; Conservative 73; Mismatches 116; Indels 13; Gaps 3;

QY 28 LEDRGLKVTITKLDPIYINVDGTSPPFGHEVFETDGAETDLDGHERFLKTKTKRN 87  
DB 1 MSSRGLKQICQKFPYINIDPGTSPYQHGSEFVTDGAETDLDGHERFLVNLKAS 60  
QY 88 NPTGQVYEQVLRNRKGDYLGATVQVPHITDEIKRKYESA--EKNDAVIEVGTVG 145  
DB 61 NVTKGIWYSISKRRGKGYGVQVPHITDEIKRKYESA--EKNDAVIEVGTVG 120  
QY 146 DIESLPELETIQVGEIGRDLPIHLTVPIYKSAEGLKTKPTQHSVLEKRTIGIOPD 205  
DB 121 DIESLPELETIQVGEIGRDLPIHLTVPIYKSAEGLKTKPTQHSVLEKRTIGIOPD 180  
QY 206 ILICRSEOPIPASERKIALFTVMKAVISAIDADTIYRIPLLAREQGLDLYVDQLRL 265  
DB 181 IIVCRTERKAPQELERKIALPCDIDPAVAVIOWDAETIYEVPLLKKEELDDIYERLKL 240  
QY 266 DVPAADLSAMEKVVGLTHTPDEVSAIVGKYVDHTDAYSLNEALIHAGIHTRHVVOIS 325  
DB 241 TCGSPDLKDMENALVHKIKNKQVTTIGLVKRYVALPANYSVASLSHAGIHNHASKIK 300  
QY 326 YIDSEITE-ABGRKALKNDAILVPGGEGGVGEGKISTVFARENKIPYLGICLQMSA 384  
DB 301 WYSADLEGAPPEPFLSDVDGIIIVPGGEGGVGEGKISALRYARENKIPYLGICLQMSA 360  
QY 385 VIEFARNVVGLEGASTFLPKSPHPVIGLITEMDAGELVTRDESDIGCTMELAQK 444  
DB 361 VVEYGRVVLGEMDANSEFSETTHPIYIDLPQKQV-----EKGQTRGLQIWA 410  
QY 445 CRLKADSLAQOLYQKQVITERRHRYEFNNQY LKQLRAAGKPSGKSLDGLVLEIIELE 502  
DB 411 CRVVKGTWDAVQDEYVERHRRHRYEFNNNYEALIEKAGATISGSPDGLVLEVAL 468

Search completed: January 29, 2004, 15:54:12  
Job time : 42.5492 secs

GenCore version 5.1.6  
Copyright (c) 1993 - 2004 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: January 29, 2004, 15:40:33 ; Search time 29.0679 Seconds  
(without alignments)  
1736.455 Million cell updates/sec

Title: US-09-941-947a-18  
Perfect score: 1623  
Sequence: 1 MOIVANPRGFCAGVDRAIE.....GIEKVFSPRELKGMQA 318

Scoring table: BLOSUM62  
Gapop 10.0 , Gapext 0.5

Searched: 1107863 seqs, 158726573 residues

Total number of hits satisfying chosen parameters: 1107863

Minimum DB seq length: 0  
Maximum DB seq length: 2000000000

Post-Processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Database :

A\_Geneseq\_19jun03.\*  
1: /SIDS1/gcgdata/geneseq/geneseq-emb1/AA1980.DAT.\*  
2: /SIDS1/gcgdata/geneseq/geneseq-emb1/AA1981.DAT.\*  
3: /SIDS1/gcgdata/geneseq/geneseq-emb1/AA1982.DAT.\*  
4: /SIDS1/gcgdata/geneseq/geneseq-emb1/AA1983.DAT.\*  
5: /SIDS1/gcgdata/geneseq/geneseq-emb1/AA1984.DAT.\*  
6: /SIDS1/gcgdata/geneseq/geneseq-emb1/AA1985.DAT.\*  
7: /SIDS1/gcgdata/geneseq/geneseq-emb1/AA1986.DAT.\*  
8: /SIDS1/gcgdata/geneseq/geneseq-emb1/AA1987.DAT.\*  
9: /SIDS1/gcgdata/geneseq/geneseq-emb1/AA1988.DAT.\*  
10: /SIDS1/gcgdata/geneseq/geneseq-emb1/AA1989.DAT.\*  
11: /SIDS1/gcgdata/geneseq/geneseq-emb1/AA1990.DAT.\*  
12: /SIDS1/gcgdata/geneseq/geneseq-emb1/AA1991.DAT.\*  
13: /SIDS1/gcgdata/geneseq/geneseq-emb1/AA1992.DAT.\*  
14: /SIDS1/gcgdata/geneseq/geneseq-emb1/AA1993.DAT.\*  
15: /SIDS1/gcgdata/geneseq/geneseq-emb1/AA1994.DAT.\*  
16: /SIDS1/gcgdata/geneseq/geneseq-emb1/AA1995.DAT.\*  
17: /SIDS1/gcgdata/geneseq/geneseq-emb1/AA1996.DAT.\*  
18: /SIDS1/gcgdata/geneseq/geneseq-emb1/AA1997.DAT.\*  
19: /SIDS1/gcgdata/geneseq/geneseq-emb1/AA1998.DAT.\*  
20: /SIDS1/gcgdata/geneseq/geneseq-emb1/AA1999.DAT.\*  
21: /SIDS1/gcgdata/geneseq/geneseq-emb1/AA2000.DAT.\*  
22: /SIDS1/gcgdata/geneseq/geneseq-emb1/AA2001.DAT.\*  
23: /SIDS1/gcgdata/geneseq/geneseq-emb1/AA2002.DAT.\*  
24: /SIDS1/gcgdata/geneseq/geneseq-emb1/AA2003.DAT.\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

# SUMMARIES

Result No.	Score	Query Match	Length	DB	ID	Description
1	1623	100.0	318	23	AAE22307	Methylenomonas 16a s
2	1623	100.0	318	23	AAU80334	Methylenomonas 16a O
3	996	61.4	316	24	ABP60423	Escherichia coli L
4	975	60.1	322	24	ABP77632	N. gonorrhoeae am
5	969.5	59.7	349	20	AAV23260	Amino acid sequenc
6	777	47.9	350	22	AAU84857	Propionibacterium
7	755	46.5	325	22	AAU90889	C. glutamicum prote
8	755	46.5	325	22	AAU78920	C. glutamicum SRT
9	743.5	45.8	329	22	AAU81220	Mycobacterium tube

10	733	45.2	335	22	AAU81137	Mycobacterium tube
11	663	40.9	310	20	AAU75672	Chlamydia pneumoniae
12	663	40.9	310	23	ABP62010	C. pneumoniae BVH
13	663	40.9	310	23	AAU09434	Chlamydia pneumoniae
14	658	40.5	352	23	ABP66088	Bifidobacterium 10
15	657	40.5	307	20	AAV37259	Amino acid sequenc
16	634.5	39.1	293	22	AAU67955	Propionibacterium
17	598.5	36.9	539	22	AAU51849	P. faecium isop
18	443.5	27.3	331	23	ABU47969	Listeria monocytog
19	305	18.8	464	21	AAU15159	A. palaeolus LTB
20	300	18.5	380	21	AAU26263	Arabidopsis thalia
21	300	18.5	380	21	AAU38001	Arabidopsis thalia
22	300	18.5	414	21	AAU26262	Arabidopsis thalia
23	300	18.5	414	21	AAU38000	Arabidopsis thalia
24	300	18.5	466	21	AAU26261	Arabidopsis thalia
25	300	18.5	466	21	AAU37999	Arabidopsis thalia
26	259.5	16.0	146	22	ABU1839	Novel human diagno
27	244.5	15.1	197	23	ABU51526	Novel human diagno
28	230	14.2	452	23	ABP93369	Helicobacter pylori
29	194.5	12.0	192	18	AAU20838	Arabidopsis thalia
30	192.5	11.9	182	18	AAU20153	H. pylori inner me
31	115	7.1	1466	22	ABU18175	Novel human diagno
32	115	7.1	2458	22	ABU24622	Novel human diagno
33	112	6.9	1033	23	ABP66122	Bifidobacterium 10
34	110	6.8	500	23	ABU48818	Listeria monocytog
35	109	6.7	578	22	AAU81233	Mycobacterium tube
36	108.5	6.7	544	12	AAU13337	HypB protein. Chl
37	104.5	6.4	539	22	AAU1848	S. epidermidis ope
38	104	6.4	488	22	AAU26267	S. epidermidis ope
39	104	6.4	494	23	ABP40179	Staphylococcus epi
40	104	6.4	454	20	AAU33930	Consensus mino aci
41	103	6.3	433	22	AAU35014	Enterococcus faeca
42	103	6.3	553	23	ABP73501	Candida albicans e
43	103	6.3	585	18	AAU26665	Yeast checkpoint c
44	103	6.3	585	20	AAU73895	Yeast MRC3 protein
45	103	6.3	896	22	AAU61465	Propionibacterium

# ALIGNMENTS

RESULT 1	AAE22307
ID	AAE22307 standard; Protein; 318 AA.
AC	AAE22307;
XX	
AC	AAE22307;
XX	
25-JUL-2002	(first entry)
XX	
DE	Methylenomonas 16a sp. dimethylallyl diphosphate (lytB) enzyme.
XX	
KW	Carotenoid; isopentenyl pyrophosphate; anthraxanthin; diet;
KW	anti-oxidant; steroid; flavour; fragrance; electro-optic application;
KW	agriculture; enzyme; dimethylallyl diphosphate; lytB.
OS	Methylenomonas 16a sp.
XX	
FN	W0200218617-A2.
XX	
PD	07-MAR-2002.
XX	
PP	04-SEP-2001; 2001MO-US27420.
XX	
PR	01-SEP-2000; 2000US-229858P.
XX	
PR	01-SEP-2000; 2000US-229907P.
XX	
PA	(DUPO) DU PONT DE NEMOURS & CO E I.
XX	
PI	Brosowski PC, Cheng O, Dicosimo DJ, Koffas M, Miller BS;
XX	Odum WJ, Picataggio SK, Roudiere PE;
XX	WPI; 2002-351711/38.
DR	N-PSDB; AAD35505.

XX Producing carotenoid compounds e.g. anthraxanthin and astaxanthin, by  
PT using microorganisms having a nucleic acid molecule encoding enzymes in  
PT the carotenoid biosynthetic pathway and which metabolize single carbon  
PT substrates

PS Claim 47; Page 123-124; 156pp; English.

XX The invention relates to a method for producing carotenoid compounds.  
CC The method comprises a transformed metabolizing host cell, comprising  
CC suitable levels of isopentenyl pyrophosphate and a nucleic acid molecule  
CC encoding an enzyme in the carotenoid biosynthetic pathway, under the  
CC control of regulatory sequences, and contacting the host cell with carbon  
CC substrate to produce a carotenoid compound. The method is useful for  
CC producing carotenoid compounds such as anthraxanthin and astaxanthin, by  
CC using microorganisms having a nucleic acid molecule encoding enzymes in  
CC the carotenoid biosynthetic pathway and which metabolize single carbon  
CC substrates. The carotenoids have potent anti-oxidant properties useful in  
CC diet, and aquaculture elements. The carotenoids are also useful as  
CC intermediates in the synthesis of steroids flavours and fragrances and  
CC compounds for potential electro-optic applications. The present sequence  
CC is Methylomonas 16a sp. dimethylallyl diphosphate (lytB) enzyme used  
CC in the invention.

XX Sequence 318 AA;

SQ Query Match 100.0%; Score 1623; DB 23; Length 318;

Best Local Similarity 100.0%; Pred. No. 2,3e-141; Mismatches 0; Indels 0; Gaps 0;

Matches 318; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 MOIVLANPFGCAGVDRAIEIVDQALBAPGAPIVRHEVHNRTVVDGLKQKAVFIEEL 60  
DB 1 MOIVLANPFGCAGVDRAIEIVDQALBAPGAPIVRHEVHNRTVVDGLKQKAVFIEEL 60  
QY 61 SDVPVGSYLIFSAGVSKVQQAERQTLVDPATCPVTKVMQYAKAKQGREVILIG 120  
DB 61 SDVPVGSYLIFSAGVSKVQQAERQTLVDPATCPVTKVMQYAKAKQGREVILIG 120  
QY 121 HAGHEVEGTMGQYEKTEGGGIIYVETPEDVRLNKKVNNPDLAVYTQTLSTMTDKMW 180  
DB 121 HAGHEVEGTMGQYEKTEGGGIIYVETPEDVRLNKKVNNPDLAVYTQTLSTMTDKMW 180  
QY 181 DALREQPPSIKEOKKDCIYATONRQDAVHDLAKISDLILVGSPPNSNRLREIAVOL 240  
DB 181 DALREQPPSIKEOKKDCIYATONRQDAVHDLAKISDLILVGSPPNSNRLREIAVOL 240  
QY 241 GKPAYLIDTYQDILKQDMLEGEIEVGVTAGASAPVAVQVVIDQLKAWGGETTSVRENSGI 300  
DB 241 GKPAYLIDTYQDILKQDMLEGEIEVGVTAGASAPVAVQVVIDQLKAWGGETTSVRENSGI 300  
QY 301 EEKVVFSPKELKKMOA 318  
DB 301 EEKVVFSPKELKKMOA 318

RESULT 2

AAU80334 ID AAU80334 standard; Protein; 318 AA.

AAU80334; 15-JUL-2002 (first entry)

XX Methylomonas 16a ORF10 lytB protein sequence.

XX Isoprenoid biosynthetic enzyme; isoprenoid compound; feed additive;  
KM kerateneid; pigment; flavour; fragrance; open reading frame 10; ORF10;  
XX lytB; IPP; isopentenyl diphosphate; dimethylallyl diphosphate.

XX Methylomonas sp.

XX WO200220733-A2.

PD 14-MAR-2002.

XX 29-AUG-2001; 2001WO-US26852.

XX 01-SEP-2000; 2000US-229907P.

XX (DUFO) DU PONT DE NEMOURS & CO E. I.

XX Cheng Q, Koffas M, Norton KC, Odom JM, Picatagallo SK, Rouviere PE;  
PI Schenle A, Tomb J;

XX WPI; 2002-383051/41.

XX DR N-PSDB; ABR50094.

XX Novel nucleic acid molecule encoding a isoprenoid biosynthetic enzyme,  
PT isolated from Methylomonas 16a, useful for the production of isoprenoid  
PT compounds

PS Claim 4; Page 84; 84pp; English.

XX The present invention relates to a new nucleic acid molecule encoding  
CC an isoprenoid biosynthetic enzyme isolated from Methylomonas 16a.  
CC The invention is useful for obtaining a nucleic acid molecule  
CC encoding an isoprenoid compound biosynthetic enzyme, and for the  
CC microbial production of isoprenoid compounds. The molecules of the  
CC organism are also useful for regulating isoprenoid biosynthesis in an  
CC organism and for producing recombinant organisms for producing various  
CC isoprenoid compounds. The nucleic acid is also useful for feed additive,  
CC for the production of kerateneid and their derivatives, isoprenoid  
CC intermediates, and as pure products useful as pigments, flavours and  
CC fragrances. The present amino acid sequence represents the Methylomonas  
CC 16a open reading frame 10 (ORF10) lytB protein of the invention, as  
CC described above. This sequence functions in the formation of IPP  
CC (isopentenyl diphosphate) and dimethylallyl diphosphate.

XX Sequence 318 AA;

SQ Query Match 100.0%; Score 1623; DB 23; Length 318;

Best Local Similarity 100.0%; Pred. No. 2,3e-141; Mismatches 0; Indels 0; Gaps 0;

QY 1 MOIVLANPFGCAGVDRAIEIVDQALBAPGAPIVRHEVHNRTVVDGLKQKAVFIEEL 60  
DB 1 MOIVLANPFGCAGVDRAIEIVDQALBAPGAPIVRHEVHNRTVVDGLKQKAVFIEEL 60  
QY 61 SDVPVGSYLIFSAGVSKVQQAERQTLVDPATCPVTKVMQYAKAKQGREVILIG 120  
DB 61 SDVPVGSYLIFSAGVSKVQQAERQTLVDPATCPVTKVMQYAKAKQGREVILIG 120  
QY 121 HAGHEVEGTMGQYEKTEGGGIIYVETPEDVRLNKKVNNPDLAVYTQTLSTMTDKMW 180  
DB 121 HAGHEVEGTMGQYEKTEGGGIIYVETPEDVRLNKKVNNPDLAVYTQTLSTMTDKMW 180  
QY 181 DALREQPPSIKEOKKDCIYATONRQDAVHDLAKISDLILVGSPPNSNRLREIAVOL 240  
DB 181 DALREQPPSIKEOKKDCIYATONRQDAVHDLAKISDLILVGSPPNSNRLREIAVOL 240  
QY 241 GKPAYLIDTYQDILKQDMLEGEIEVGVTAGASAPVAVQVVIDQLKAWGGETTSVRENSGI 300  
DB 241 GKPAYLIDTYQDILKQDMLEGEIEVGVTAGASAPVAVQVVIDQLKAWGGETTSVRENSGI 300  
QY 301 EEKVVFSPKELKKMOA 318  
DB 301 EEKVVFSPKELKKMOA 318

RESULT 3

ABP60423 ID ABP60423 standard; Protein; 316 AA.

XX ABP60423;

XX 31-MAR-2003 (first entry)

XX Escherichia coli LytB protein SEQ ID NO 2.  
DE  
XX Escherichia coli; mevalonate-independent isoprenoid synthesis pathway;  
XX MEK-way; Gcgb; LytB; T cell; asthma; Crohn's disease; multiple sclerosis;  
XX autoimmune disease; osteoporosis; tumour; antisthmatic; antilucer;  
XX antiinflammatory; neuroprotective; immunosuppressive; antiallergic;  
XX osteopathic; enzyme.  
XX  
OS Escherichia coli.  
XX  
PN DE10119905-A1.  
XX  
PD 24-OCT-2002.  
XX  
PF 23-APR-2001; 2001DE-1019905.  
XX  
PR 23-APR-2001; 2001DE-1019905.  
XX  
PA (JOMA-) JOMAA PHARMAKA GMBH.  
XX  
XI Jomaa H, Eberl M, Altincicek B;  
XX  
XX WPI; 2003-113392/11.  
DR N-PSDB; AM59204.  
XX  
XX Enriching intermediates in the mevalonate-independent pathway of  
PT isoprenoid synthesis, useful for therapeutic activation of T cells,  
PT comprises altering enzymatic activity in the pathway -  
XX  
PS Disclosure; Page 6-8; 10pp; German.

	Query March Best Local Matches 192;	Similarity 61.3%; Conservative 59;	Score 986; Pred. No. 1.8e-83; Mismatches 58;	DB 24; Length 316; Indels 4; Gaps 2
QY	1	MOVLANRGEFCAGVDRALEIVDQAEAFGAPIVYRHHVHNRITVVDGKQGAFFIEL	60	
DB	1	MOVLANRGEFCAGVDRALEIVDQAEAFGAPIVYRHHVHNRITVVDGKQGAFFIEL	60	
QY	61	SDVAVGSLISAHCVSKVQQAERQLVTPDACPLVTKRMQYAGHAKGRRVILIG	120	
DB	61	SEVDDGALLISAHCVSQAVRNEASRLVTFDACPLVTKRMVAVARRSGESILIG	120	
QY	121	HAGHPEVEGTMGQYEKTEGGGIYLVETPEDEVRYLNKNPNVDLAVYTQTLMTFTKVV	180	
DB	121	HAGHPEVEGTMGQYS--NPEGGMVTVESPDPVWKLTVANGEXLSPMTQTLTSDVTSVY	178	
QY	181	DALHEQFSTIEKKKKDICYATONRQDPAVHDLAKLSLLIVVGSVNSNSNRLEPIANQL	240	
DB	179	DALKRRPFKVIQPRKDDICYATNNQKQAVRALAEQAEVLLVVGSSNNSNNLALAEQAM	238	
QY	241	GKPAVLEIDTYODLQKQWLEGIIVWGVTVAGASAPVLEQVVIDQLKAMGETTVRENGSI	300	

```

Db      239  GKAFILIDAKIOIBWWEKVCVGTAGASAPDILVNVARLOOLGGGAIPLE--GR 296
QY      301  EEKVFESIPKEIK 313
        ||:||||:
Db      297  EENIVFEVPEKLR 309

```

#### RESULT 4

ID ABP77632 standard; Protein; 322 AA.

AC ABP77632;

DT 07-MAR-2003 (first entry)

DE N. gonorrhoeae amino acid sequence SEQ ID 1794.

**KW** Antibacterial; infection; vaccine; gene therapy.

05 *Neisseria gonorrhoeae*.

PN MO200279243-A2.

PD 10-OCT-2002.

PF 12-FEB-2002; 2002WO-IB02069.  
XX

PR 12-FEB-2001; 2001GB-0003424.  
YY

PA (CHIR-) CHIRON SPA.  
 Y Y

PI Fontana MR, Pizza M, Masignani V, Monaci E, XY

DR WPI; 2003-058415/05  
DR N-PSTB; ABZ38602

XX New protein from  
PT

PT medicament for treating or preventing

PS Disclosure; page 31 /; 815pp; English: XX

Also disclosed are the nucleic acid molecules encoding the proteins and

comprising the protein, nucleic acid or antibody is useful for the

infection, this may be in the form of a vaccine or gene therapy.

sequences given in records ADF/6/50-ADF0149 represent molecules of the invention.

Seq	Sequence	322 AA
1	...	...
2	...	...
3	...	...
4	...	...
5	...	...
6	...	...
7	...	...
8	...	...
9	...	...
10	...	...
11	...	...
12	...	...
13	...	...
14	...	...
15	...	...
16	...	...
17	...	...
18	...	...
19	...	...
20	...	...
21	...	...
22	...	...
23	...	...
24	...	...
25	...	...
26	...	...
27	...	...
28	...	...
29	...	...
30	...	...
31	...	...
32	...	...
33	...	...
34	...	...
35	...	...
36	...	...
37	...	...
38	...	...
39	...	...
40	...	...
41	...	...
42	...	...
43	...	...
44	...	...
45	...	...
46	...	...
47	...	...
48	...	...
49	...	...
50	...	...
51	...	...
52	...	...
53	...	...
54	...	...
55	...	...
56	...	...
57	...	...
58	...	...
59	...	...
60	...	...
61	...	...
62	...	...
63	...	...
64	...	...
65	...	...
66	...	...
67	...	...
68	...	...
69	...	...
70	...	...
71	...	...
72	...	...
73	...	...
74	...	...
75	...	...
76	...	...
77	...	...
78	...	...
79	...	...
80	...	...
81	...	...
82	...	...
83	...	...
84	...	...
85	...	...
86	...	...
87	...	...
88	...	...
89	...	...
90	...	...
91	...	...
92	...	...
93	...	...
94	...	...
95	...	...
96	...	...
97	...	...
98	...	...
99	...	...
100	...	...

Query Match	60.18; Score 975; DB 24; length 322;
-------------	--------------------------------------

Matches 192; Conservative 48; Mismatches 66; Indels 6; Gaps 2;

QY 3 IVLANPRGFCAGVDRAIEIVDOAIEAFGAPIYVRHEVTVHNRVTVDGLKÖKGAVFIEELSD 62

Db 6 IILANPRGFCAGVDRAISIVERAL EEFAGPVYVRHEVHNK FVDNLRKGA VFIEDLAE 655

63 VPVGSYLIFSAHGVSKVEVQQAEBRQLTVFDATCP LVTKVHMQVAKHAKQGREVILGHA 122

66 VPPGATLVYSAHGVS KAVQGEAERGFRVFDATCP LVTKVHKEVARLDAQDCEIIMIGHK 12

QY 123 GHPEVBTMGQYEKCTBGGGIYLV3TPEDEVRLKVNPNPDLAYVTQFTLSMTDTKMWDA 18

Db 126 GHAEVEGTGQLA P - - - GKMLLVETVGDVAKLEVRNPDKLAYVSQTTLSVDETKDIIAA 18

183 LREQPSIKKKDDICYATQNRQDAVHDLAKISDLILVVGSPNSSNSNRLREIAVQLGK 24

Db 182 LNARFPIRNPHKEDICYATTNRQTAVKELAEQCDIVIVGSPNSSNSNRLREVAASRGI 24

QY 243 PAYLIDVTQDLKODMLEGIEVVGVTAGASAPVLTVOEVIDOKAMGGFTTSVRENSGIE 302  
 DB 242 DAYMDNABYIQRTPEGSKSVGTAGASAPVLTVEVLAIRGMSHET--VREGGABE 299  
 QY 303 KVFESIPKELKK 314  
 DB 300 SIVFVLPEKELRR 311

## RESULT 5

AAI29260  
 ID AAY29260 standard; Protein; 349 AA.  
 AC AAY29260;

DT 25-OCT-1999 (first entry)

DE Amino acid sequence of a virulence factor encoded by ORF4234c.

KM Human pathogen; virulence polypeptide; virulence factor;  
 XX pathogenic infection; Pseudomonas aeruginosa infection.

OS Pseudomonas aeruginosa.

PN WO9927129-A1.

PD 03-JUN-1999.

PF 25-NOV-1998; 98WC-US25247.

PR 25-NOV-1997; 97US-0066517.

PA (GENO) GEN HOSPITAL CORP.

PI Ausubel F, Cao H, Drenkard E, Goodman HM, Mahajan-Miklos S;  
 PI Rahne IG, Tan M, Tsongalis J;

DR WPI; 1999-357851/30.

XX Virulence factors useful in developing disease treatments

PS Disclosure; Fig 4; 228pp; English.

CC The present sequence represents a Pseudomonas aeruginosa polypeptide  
 CC sequence. P. aeruginosa is an opportunistic human pathogen present in  
 CC soil water and plants. The specification describes virulence polypeptides  
 CC and nucleic acid sequence encoding such polypeptides. These sequences  
 CC can be used to identify a compound which is capable of decreasing the  
 CC expression of a pathogenic virulence factor. Compounds that inhibit  
 CC the expression or activity of virulence factor polypeptides can be  
 CC used to treat pathogenic infections, especially where the infection  
 CC is a P. aeruginosa infection.  
 CC note: the sequences given in the specification were poorly legible, and  
 CC in some instances assumptions were made as to the identity of the  
 CC residue; it is therefore possible that the sequence given below is  
 CC not entirely correct.

XX Sequence 349 AA;

Query Match 59.7%; Score 969.5; DB 20; Length 349;  
 Best Local Similarity 61.0%; Pred. No. 5.9e-81;  
 Matches 191; Conservative 47; Mismatches 72; Indels 3; Gaps 2;

QY 1 MGIIVANRPGCAGYDRAIRIYDQAIEMRGAPIVRHEVYNNRTVYDGAKQKAFIEEL 60  
 DB 36 MGIKAMPFGFCAGYDRAIRIYDQAIEMRGAPIVRHEVYNNRTVYDGAKQKAFIEEL 95  
 QY 61 SDVPVSYLIFSAHGSVSKVQDAEERQTVDPATCPVTKVQVQVAKAKQGREVILIG 120  
 DB 96 DQVPMNVITVIFSAHGSVSKVQDAEERQTVDPATCPVTKVQVQVAKAKQGREVILIG 155  
 QY 121 HAGHEVBEITMQYKCTBGGGIYVETPEYVRNKKVNNPNDLAVYQTTSMDTKMV 180

DB 156 HEGHEVBEITMQYD-ASNGGAIYLVDEADVAALBYRKEPAHYVYQTTSMDTSKVI 214  
 QY 181 DALRQPFSTIEQKKDDICVATQNRQDANFDLACISLILVSGSPNSNSRLAEIAYQL 240  
 DB 215 DALRQPFQIQPFKMDICVATQNRQDANFDLACISLILVSGSPNSNSRLAEIAYQL 274  
 QY 241 GKPAVLIDVTQDLKODMLEGIEVVGVTAGASAPVLTVOEVIDOKAMGGFTTSVRENSGIE 300  
 DB 275 GTPAVLIDGAEADMQRGFCGVRIRIGITRAGASAPVLTVRGVIQLRENG--ASEGQELBGR 332  
 QY 301 EKVVFESIPKELK 313  
 DB 333 EKVVFESIPKELR 345

## RESULT 6

AAU58857  
 ID AAU58857 standard; Protein; 350 AA.

AC AAU58857;

DT 13-FEB-2002 (first entry)

DE Propionibacterium acnes immunogenic protein #19753.

KM SAPHO syndrome; synovitis; acne; pustulosis; hypertosis; osteomyelitis;  
 XX uveitis; endophthalmitis; bone; joint; central nervous system; ELISA;  
 XX inflammatory lesion; acne vulgaris; enzyme linked immunosorbent assay;  
 XX dermatological; osteopathic; neuroprotectant.

OS Propionibacterium acnes.

PN WO200181581-A2.

PD 01-NOV-2001.

PF 20-APR-2001; 2001WO-US12865.

PR 21-APR-2000; 2000US-199047P.

PR 02-JUN-2000; 2000US-208841P.

PR 07-JUL-2000; 2000US-216747P.

PA (CORI-) CORIXA CORP.

PI Skeiky YAM, Persing DH, Mitcham JL, Wang SS, Bhadia A;  
 PI L'walbomeuve J, Zhang Y, den S, Carter D;

DR WPI; 2001-616774/71.

XX N-PSDB; AAS59596.

XX Propionibacterium acnes polypeptides and nucleic acids useful for  
 XX vaccinating against and diagnosing infections, especially useful for  
 XX treating acne vulgaris -

PS Example 1; SEQ ID No 20052; 1069pp; English.

CC Sequences AAU39105-AAU588017 represent Propionibacterium acnes immunogenic  
 CC polypeptides. The proteins and their associated DNA sequences are used in  
 CC the treatment, prevention and diagnosis of medical conditions caused by  
 CC P. acnes. The disorders include SAPHO syndrome (synovitis, acne,  
 CC pustulosis, hypertosis and osteomyelitis), uveitis and endophthalmitis.  
 CC P. acnes is also involved in infections of bone, joints and the central  
 CC nervous system, however it is particularly involved in the inflammatory  
 CC lesions associated with acne vulgaris. A method for detecting the  
 CC presence or absence of P. acnes in a patient comprises contacting a  
 CC sample with a binding agent that binds to the proteins of the invention  
 CC and determining the amount of bound protein in the sample. The  
 CC polypeptides may be used as antigens in the production of antibodies  
 CC specific for P. acnes proteins. These antibodies can be used to  
 CC downregulate expression and activity of P. acnes polypeptides and  
 CC therefore treat P. acnes infections. The antibodies may also be used as  
 CC diagnostic agents for determining P. acnes presence, for example, by  
 CC enzyme linked immunosorbent assay (ELISA).

CC Note: The sequence data for this patent did not form part of the printed specification, but was obtained in electronic format directly from WIPO  
 CC at ftp.wipo.int/pub/published\_pct\_sequences.

XX Sequence 350 AA;

Query Match 47.9%; Score 777; DB 22; Length 350;  
 Best Local Similarity 48.1%; Pred. No. 3,56-63;  
 Matches 152; Conservative 68; Mismatches 84; Indels 12; Gaps 4;

QY 2 QIVLANPFGCAGVDRAIEIVDQIAEAFPIYHEVHNFTVVDLKKQKAVFIEELS 61  
 DB KLVIAAPRGYCAVDRAIVETVERKALIEYGAIVYRKQIVNRRHVETLEGRGAIPIVDELD 97  
 QY 62 DVPGSYLIFSAGVSKVEVQOAEERQLTFPDATCELVTKYMNQVAKHAKQGRVILLGH 121  
 DB EVDPAIVSVSAGVSKVEVQOAEERQLTFPDATCELVTKYMNQVAKHAKQGRVILLGH 157

QY 122 AGHPEVETGQYKCTEGGGIYLVETPEDVRLKVNPNDLAVYVQTTLSTMTDKVMVD 181  
 DB AGHPEVETGQYKCTEGGGIYLVETPEDVRLKVNPNDLAVYVQTTLSTMTDKVMVD 212

QY 182 ALABQPSISKEQKDDICVATQNRQDAVHDLAKISDLIVVSGPSNSNSRLKEIYVQIG 241  
 DB RIRDLHPQLIDPPSDICVATQNRQDAVHDLAKISDLIVVSGPSNSNSRLKEIYVQIG 272

QY 242 -KRAYLIDTQDKOMPLIEGVVGTAGASAEVLYQVEYIDOL--KAMGETTSVRENS 298  
 DB AKASVRYDANGLIEETWLDGVSTIGTSGSVPEALVQSYIDILISKGM---PPAEET 328

QY 299 GIEEKVVFSTPEKELK 314  
 DB 329 LIEESLIFALPQLRK 344

# RESULT 7

AAG90889 standard; Protein; 325 AA.

XX AAG90889;

DT 26-SEP-2001 (first entry)

DE C glutamicum protein fragment SEQ ID NO: 4643.

KM Corynebacterium; amino acid synthesis; vitamin; saccharide;  
 KM organic acid synthesis.

OS Corynebacterium glutamicum.

PN EPI108790-A2.

PD 20-JUN-2001.

PF 18-DEC-2000; 2000EP-0127688.

PR 16-DEC-1999; 99UP-0377484.

PR 07-APR-2000; 2000UP-0159162.

PR 03-APR-2000; 2000UP-0280988.

PA (KYOW) KYOWA HAKKO KOGYO KK.

XX Nakagawa S, Mizoguchi H, Ando S, Hayashi M, Ochiai K, Yokoi H;  
 PI Tateishi N, Senoh A, Ikeda M, Ozaki A;

DR WPI; 2001-376931/40.

DR N-PSDB; AAH66108.

PT Novel polynucleotides derived from Corynebacterium bacteria, for identifying  
 PT mutation point of a gene, measuring expression of a gene, analysing  
 PT expression profile or pattern of a gene and identifying homologous gene

PS Claim 17; SEQ ID NO: 4643; 246pp + Sequence Listing; English.

XX The present invention provides a number of nucleotide and protein  
 CC sequences from the Corynebacterium bacteria Corynebacterium glutamicum. These  
 CC are useful for identifying the mutation point of a gene derived from a  
 CC mutant of corynebacterium, measuring expression amount and  
 CC analysing the expression profile or expression pattern of a gene derived  
 CC from Corynebacterium bacteria, and identifying a homologue of a gene derived  
 CC from Corynebacterium bacteria. Corynebacterium bacteria are useful for producing  
 CC amino acids, nucleic acids, vitamins, saccharides and organic acids,  
 CC particularly L-lysine. The present sequence is a protein described  
 CC in the exemplification of the invention.

CC Note: The sequence data for this patent did not form part of the printed  
 CC specification, but was obtained in electronic format directly from the  
 CC European Patent Office.

XX Sequence 325 AA;

Query Match 46.5%; Score 755; DB 22; Length 325;  
 Best Local Similarity 48.3%; Pred. No. 3,46-61;  
 Matches 152; Conservative 67; Mismatches 86; Indels 10; Gaps 4;

QY 2 QIVLANPFGCAGVDRAIEIVDQIAEAFPIYHEVHNFTVVDLKKQKAVFIEELS 61  
 DB KLVIAAPRGYCAVDRAIVETVERKALIEYGAIVYRKQIVNRRHVETLEGRGAIPIVDELD 74

QY 62 DVPGSYLIFSAGVSKVEVQOAEERQLTFPDATCELVTKYMNQVAKHAKQGRVILLGH 121  
 DB EVDPAIVSVSAGVSKVEVQOAEERQLTFPDATCELVTKYMNQVAKHAKQGRVILLGH 134

QY 122 AGHPEVETGQYKCTEGGGIYLVETPEDVRLKVNPNDLAVYVQTTLSTMTDKVMVD 179  
 DB AGHPEVETGQYKCTEGGGIYLVETPEDVRLKVNPNDLAVYVQTTLSTMTDKVMVD 189

QY 180 VDALEQPSISKEQKDDICVATQNRQDAVHDLAKISDLIVVSGPSNSNSRLKEIYVQIG 239  
 DB VRELKVPQLODPPSDICVATQNRQDAVHDLAKISDLIVVSGPSNSNSRLKEIYVQIG 249

QY 240 LGRP-AYLIDTQDKOMPLIEGVVGTAGASAEVLYQVEYIDOLKAMGETTSVRENS 298  
 DB NGADNAYLVDAVBIDPAPFESVETIGTSGSVPEALVQSYIDILAEGRYD--DVEEVT 307

QY 299 GIEEKVVFSTPEKELK 313  
 DB 308 SAEKIVFALPQLRK 322

# RESULT 8

AAB78920 standard; Protein; 325 AA.

XX AAB78920;

DT 30-APR-2001 (first entry)

DE C. glutamicum SRT protein sequence SEQ ID NO:100.

KM Corynebacterium glutamicum; stress; resistance; tolerance; SRT;  
 KM fine chemical production; organic acid; proteinogenic amino acid;  
 KM nonproteinogenic amino acid; purine base; pyrimidine base; nucleoside;  
 KM nucleotide; lipid; saturated fatty acid; unsaturated fatty acid; diol;  
 KM carbohydrate; aromatic compound; vitamin; cofactor; polypeptide; enzyme;  
 KM evolutionary study; environmental hazard; fermentation.

OS Corynebacterium glutamicum.

PN WO200100804-A2.

PD 04-JAN-2001.

PF 23-JUN-2000; 2000WO-IB00922.

PR 25-JUN-1999; 99US-0141031.

PR 01-JUL-1999; 99DE-1030429.  
 PR 01-JUL-1999; 99US-0142692.  
 PR 08-JUL-1999; 99DE-1031413.  
 PR 08-JUL-1999; 99DE-1031457.  
 PR 08-JUL-1999; 99DE-1031541.  
 PR 09-JUL-1999; 99DE-1032209.  
 PR 09-JUL-1999; 99DE-1032230.  
 PR 14-JUL-1999; 99DE-1032230.  
 PR 27-AUG-1999; 99US-1040764.  
 PR 27-AUG-1999; 99US-0151214.  
 PR 31-AUG-1999; 99DE-1041382.

XX (BAD) BASF AG.

PI Pompejus M, Kroegeer B, Schroeder H, Zelder O, Haberhauer G, Lee H;  
 PI Kim H;

XX WPI; 2001-061972/07.  
 XX N-PSDB; AAF71033.

XX New isolated *Corynebacterium glutamicum* nucleic acid encoding a stress,  
 PT tolerance or resistance protein, for production or modulation of  
 PT production of fine chemicals, such as, e.g. amino acids, lipids,  
 PT carbohydrates, or enzymes -

XX Claim 20; Page 249-250; 526pp; English.

XX AAF70984 to AAF71133 encode the *Corynebacterium glutamicum* stress,  
 CC resistance, and tolerance (SRT) proteins given in AAB78871 to AAB79020.  
 CC The C. glutamicum SRT genes (I) can be used in vectors (II) for  
 CC expression in host cells and production of fine chemicals, such as, an  
 CC organic acid, a proteinogenic or nonproteinogenic amino acid (preferred),  
 CC a purine or pyrimidine base, a nucleoside, a nucleotide, a lipid, a  
 CC saturated or unsaturated fatty acid, a diol, a carbohydrate, an aromatic  
 CC compound, a vitamin, a cofactor, a polypeptide, or an enzyme. The fine  
 CC chemical production can be modulated. The presence of (I) or the SRT  
 CC proteins (III) encoded by them are used for diagnosing the presence  
 CC or activity of *Corynebacterium diptheriae*. (I), (II), (III) and host  
 CC cells containing them can be used to map the genomes of organisms related  
 CC to C. glutamicum, to identify and localise C. glutamicum sequences of  
 CC interest, in evolutionary studies, in determination of SRT protein  
 CC regions required for function, in modulating the SRT protein activity,  
 CC and in modulating the activity of an SRT pathway. (II) are used to permit  
 CC C. glutamicum to survive in an environment that is normally  
 CC environmentally or chemically hazardous to it. (I) and protein molecules  
 CC encoded by it increase the survival of C. glutamicum to chemical and  
 CC environmental hazards and provide a means for continued growth and  
 CC multiplication in large scale fermentative growth conditions. By  
 CC increasing the growth rate or maintaining a normal growth rate in poor or  
 CC toxic conditions, the yield, production and/or efficiency or production  
 CC of fine chemicals from a culture may be increased.

XX Sequence 325 AA;

Query Match 46.5%; Score 755; DB 22; Length 325;

Best Local Similarity 48.3%; Pred. No. 3.4e-61; Mismatches 86; Indels 10; Gaps 4;

Matches 152; Conservative 67; Mismatches 86; Indels 10; Gaps 4;

QY 2 QIVANPFGCAGVDAIETVDQAIKAFGAPIVYHEVHNRRTVVDGKQGAFTIELS 61  
 DB 15 KILLAAPRGYCAGVRAIVETVRALEBYCAPTYVEKEIVHRYVYDTLAEKKAIFVNAS 74  
 QY 62 DVPVGSYLIFSAGVSKVQQAERQULTVPDQTCPLVTYKHMVQAKAKQREVTILGH 121  
 DB 75 EAPGAGNMFSAHGSFPMVHEBAKAKITADACPLVTYKHEQORPRKQGFILFIGH 134  
 QY 122 AGHPEVSGTGGYBECTGGGGLVETPEDVNLK--VNNPDLAYVTQTLLSMTDTVM 179  
 DB 135 EGHHEEGTGMGSHSVET-----HIVDGVAGIATLPEFLNDEPNLIMLSQTLLSVDTWBI 189  
 QY 180 VALAEOPFSIKQKKDVCATQNRQADVHGLAKISDILVWGSPPNSNSRLBEIAVQ 239  
 DB 190 VAELEKPEFQLODPSSDDICATQNRQAVVKAIAERCEIMLTAVGSRNSSNRYLVEVAKQ 249

QY 240 LGKP-AYCIDPTVODLKDQWLRGIEVGVAGASAPVYLVDQVINDLKAWGETTSYRENS 298  
 DB 250 NGADNAYIVDAKATDPAKPEGVETIGISSGASVPEILVQGVETRLAEFGTD--DVEEYV 307  
 QY 299 GIERKVPYSIPKEIK 313  
 DB 308 SAAEKIVPALPRVLR 322

RESULT 9

AA681220  
 ID AAG81220 standard; Protein, 329 AA.

XX AAG81220;

AC AAG81220; (first entry)

DE Mycobacterium tuberculosis potential drug target protein SEQ ID 271.

XX Drug target; growth; organism viability; characterisation.

OS Mycobacterium tuberculosis.

XX WO200135317-A1.

XX 17-MAY-2001.

XX 13-NOV-2000; 2000MO-US31152.

XX 12-NOV-1999; 99US-0165086.

XX 12-NOV-1999; 99US-0165124.

XX 01-FEB-2000; 2000US-0179531.

XX (BEGC) UNIV CALIFORNIA.

PI Eisenberg D, Rotstein SH, Marcotte EM;

DR WPI; 2001-329193/34.

DR N-PSDB; AAH52071.

PT Identifying nucleotide or polypeptide sequence for use as drug target,  
 PT involves providing algorithm that analyzes a functional relationship  
 PT between nucleotide or polypeptide sequences, and comparing the  
 PT sequences -

PS Disclosure; Page 187; 207pp; English.

XX This invention relates to a method for identifying a nucleotide or  
 CC polypeptide sequence that may be a drug target, or essential for growth  
 CC or viability of an organism. Polynucleotide sequences AAH51947 - AAH52092  
 CC represent DNA encoding proteins AAG81096 - AAG81241, Mycobacterium  
 CC tuberculosis proteins which are potential drug targets. The DNA and  
 CC protein sequences are used to illustrate the method of the invention. The  
 CC method involves providing an unknown nucleotide or polypeptide sequences,  
 CC and comparing it to a number of sequences along with at least one  
 CC algorithm capable of analysing a functional relationship between  
 CC nucleotide and polypeptide sequences. The method is useful for  
 CC characterising the function of nucleic acids and polypeptides that may be  
 CC useful as a target for a drug or essential for the growth or viability of  
 CC an organism.

XX Sequence 329 AA;

Query Match 45.8%; Score 743.5; DB 22; Length 329;

Best Local Similarity 47.6%; Pred. No. 4e-60; Mismatches 94; Indels 11; Gaps 3;

Matches 150; Conservative 60; Mismatches 94; Indels 11; Gaps 3;

QY 3 IVLANPFGCAGVDAIETVDQAIKAFGAPIVYHEVHNRRTVVDGKQGAFTIELSD 62  
 DB 20 VLASPPSPFCAGVRAIVETVRALEBYCAPTYVEKEIVHRYVYDTLAEKKAIFVNAS 79  
 QY 63 V----PVGSYLIFSAGVSKVQQAERQULTVPDQTCPLVTYKHMVQAKAKQREVTIL 118

```

Db      80  IPDPPEGAVVPSAHSVSPAVRAGADERGLGVADNCPVLAVVHAAARARQTVVF 139
Qy      119 IGHAHGEVAGTGMQVEKCTEGGCIYIVETPEEDYRNKXANPNDAVYTOTLISMTDTKY 178
Db      140 IGHAHGEETEGTGVAPRST-----LLVQPPAVAAALNIEGTQLSYLTOTTLALDETAD 194
Qy      179 MDVALREQPSIKQKQKDDICVATQNRQDAVHDLAKISDLILVVGSPNSNSNRLEIAV 238
Db      195 VIDALARFPTLQPSDEDICVATTRQRLQSGMGECDVIVIGSNSNSRRLVELAQ 254
Qy      239 QLKCPALLIDTYDQLKQDMLGIEVGVGTAGASAPETLVGEVTDLKAWGETTSVRENS 298
Db      255 RSGTPAYLLIDGPDIDIEEMLSVSTIGVTAGASAPRLVGQVTDALRGVAST--VVERS 312
Qy      299 GIEKVFESIPKELK 313
Db      313 IATEVTRFGLPKQVR 327

```

## RESULT 10

AA81137  
ID AA81137 standard; Protein; 335 AA.

AC AA81137;

DT 04-SEP-2001 (first entry)

DE Mycobacterium tuberculosis potential drug target protein SEQ ID 188.

DM Drug target; growth; organism viability; characterisation.

OS Mycobacterium tuberculosis.

PN WO200135317-A1.

PD 17-MAY-2001.

PF 13-NOV-2000; 2000MO-US31152.

PR 12-NOV-1999; 99US-0165086.

PR 12-NOV-1999; 99US-0165124.

PR 01-FEB-2000; 2000US-0179531.

PA (RBGC ) UNIV CALIFORNIA.

PI Eisenberg D, Rotstein SH, Marcotte EM;

PI WPI; 2001-329193/34.

DR N-PSDB; AAH51988.

PT Identifying nucleotide or polypeptide sequence for use as drug target,

PT involves providing algorithm that analyzes a functional relationship

PT between nucleotide or polypeptide sequences, and comparing the

PS sequences -

PS Disclosure; Page 165; 207pp; English.

CC This invention relates to a method for identifying a nucleotide or

CC polypeptide sequence that may be a drug target, or essential for growth

CC or viability of an organism. Polynucleotide sequences AAH51947 - AAH52092

CC represent DNA encoding proteins AA81096 - AA81241, Mycobacterium

CC tuberculosis proteins which are potential drug targets. The DNA and

CC protein sequences are used to illustrate the method of the invention. The

CC method involves providing an unknown nucleotide or polypeptide sequences,

CC and comparing it to a number of sequences along with at least one

CC algorithm capable of analysing a functional relationship between

CC nucleotide and polypeptide sequences. The method is useful for

CC characterizing the function of nucleic acids and polypeptides that may be

CC useful as a target for a drug or essential for the growth or viability of

CC an organism.

CC Sequence 335 AA;

```

Query Match      45.2%; Score 733; DB 22; Length 335;
Best Local Similarity 47.0%; Pred. No. 3,8e-59;
Matches 147; Conservative 66; Mismatches 92; Indels 8; Gaps 3;
Qy      2 QVLNPRFCAGVRALEIVDQALEAGAPYIVHEVVENRTVVDLQKQGAFLIEIS 61
Db      27 RLIALPEPRYCGVRAVETVRBALQKQGPYYVHEIYENRHVDTLAKAGAVVEEIE 86
Qy      62 DVPVGSYLLIPSAGVSKVQQAERQLTVPDCTPLVTKVMQVAKAKQGREVTLIGH 121
Db      87 QVPEGAIVFSAHGAPLVTHVSASERNQVIDATCPLVTKVNEARRRFRARDYDILLIGH 146
Qy      122 ACHPEVEGTMQVEKCTEGGCIYIVETPEEDYRNKLVNPNDAVYTOTLISMTDTKYWD 181
Db      147 EGHREVGTAGE-----APDHQVLVDGVADVQVTRDEKVMVLSQTTLSVDEMEIVG 201
Qy      182 ALRQPSIKQKQKDDICVATQNRQDAVHDLAKISDLILVVGSPNSNSNRLEIAVQIG 241
Db      202 RLRRRFPKQDPPSDICVATQNRQVAVKAMAPCECEIVIVGSRSSNSVRLEVALGAG 261
Qy      242 -KPAVLIDTYDQLKQDMLGIEVGVGTAGASAPETLVGEVTDLKAWGETTSVRENSGI 300
Db      262 AAPAHLVPMADDDISAMLDGVTTGVTSASVPELVGRGLERLAECCGYD--VQPVTTA 319
Qy      301 BEKVFESIPKELK 313
Db      320 NETLVFALPRELR 332

```

## RESULT 11

AA35672  
ID AA35672 standard; Protein; 310 AA.

AC AA35672;

DT 13-SEP-1999 (first entry)

DE Chlamydia pneumoniae cellular envelope protein.

DM Respiratory disease; pneumonia; bronchitis; heart disease; sarcoidosis;

DM sinusitis; purulent otitis media; erythema nodosum; pharyngitis;

DM vaccine; neutralising epitope.

OS Chlamydia pneumoniae.

PN WO9927105-A2.

PD 03-JUN-1999.

PF 20-NOV-1998; 98WO-IB01890.

PR 04-NOV-1998; 98US-0107078.

PR 21-NOV-1997; 97FR-0014673.

PA (GBST ) GENSET.

PI Griffais R;

PI WPI; 1999-357842/30.

PT Genome sequence of Chlamydia pneumoniae

PT Page 1389; Disclosure; 1912pp; English.

CC AA35672-1389 represent the proteins encoded by all the open reading

CC frames in the complete genome (see AA35672-1389) of Chlamydia pneumoniae.

CC C. pneumoniae causes respiratory diseases such as pneumonia and

CC bronchitis and is thought to be a contributing factor in heart

CC disease, sarcoidosis, sinusitis, purulent otitis media, erythema

CC nodosum or pharyngitis. The polypeptides encoded by the open reading

CC frames of the C. pneumoniae genome (see AA35672-1389) can be used in

CC immunogenic compositions as vaccines. Vectors containing C. pneumoniae

CC nucleotides sequences can also be used as immunogenic compositions,  
CC especially where the vector directs the expression of a neutralising  
CC epitope of C. pneumoniae.

XX Sequence 310 AA;

Query Match 40.9%; Score 663; DB 20; Length 310;

Best Local Similarity 45.0%; Pred. No. 9.9e-53;  
Matches 143; Conservative 59; Mismatches 98; Indels 18; Gaps 3;

CC 2 QIVLAMPKPGCAGVDRALEIVDOALTEAFGAPVYRHEVYHNRVTVDGKQKGAFTIEELS 61

DB 3 KLILCNPRGPGSGVVRALQVVEVLEKMGAPVYVKEHIVNHRVYVVALAKAGALFVEELV 62

QY 62 DVPVGSYLIFSAHAGSKKEVQGEAEEROLTVFADTCPLVTXHMVAKHAKQGREVILIGH 121

DB 63 DVEGERVIYSAGHGPSVRARAKRLIDIDATCGLVTKHSAKLYASRGYKILILIGH 122

QY 122 AGHPEVEGTMQYKTEGGGIYLVETPEDEVRLKVNPNDAVYVQTLLSMTDTKAWD 181

DB 123 KKHVEVIGIVEVPE-----HITVEKVADEVALPFSSDPLFYITQTLLSDVQGEISS 177

QY 182 ALREQPSIIEQKDDICATQNRQDAVHDLAKISDLILVYSGPSNSNRLREIAYQLG 241

DB 178 ALTKRPSIITLPSSICATITNRQKALRSVLSRVNYVYVGVNNSNRLREVALRGG 237

QY 242 KPAVLIIDTYODLKQDMLGIEVGVGTAGASAPVAVQEVLDQKAMGETTSVRANSGLIE 301

DB 238 VPADLIINNPEDIDITNVHSGDIAMTAGASTPEDVVOACIRKL-----SSLIGLQVE 290

QY 302 -----EKVYFSPKELK 313

DB 291 NDIFAVEDVFOLPKELR 308

RESULT 12

ABP62010

ID ABP62010 standard; Protein; 310 AA.

AC ABP62010;

DT 08-OCT-2002 (first entry)

XX C. pneumoniae BVH-CPN18 from strain CWL-029.

XX CWL-029; BVH-CPN; antibiotic; antibacterial; Chlamydial infection;

XX sinusitis; pharyngitis; bronchitis; pneumonia; asthmatic bronchitis;

XX adult-onset asthma; chronic obstructive pulmonary disease; COPD;

XX atherosclerosis;

XX Chlamydia pneumoniae.

XX EPI219635-A2.

XX 03-JUL-2002.

XX 21-DEC-2001; 2001EP-0130295.

XX 21-DEC-2000; 2000US-256941P.

XX (SHIR-) SHIRE BIOCHEM INC.

XX Couture F, Hamel J, Brodeur BR, Martin D;

XX WPI; 2002-530680/57.

XX N-PSDB; ABQ2504.

XX New Chlamydia pneumoniae proteins or antigens, useful for the

XX prophylactic or therapeutic treatment of Chlamydial bacterial

XX infections, e.g. sinusitis, pharyngitis, bronchitis, or chronic

XX obstructive pulmonary disease

XX Claim 9; Figure 36; 122pp; English.

XX The invention relates to novel isolated polypeptides from Chlamydia

CC pneumoniae. The polypeptides of the invention have antibiotic and

CC antibacterial activity. The polypeptide or composition of the invention

CC is useful for the prophylactic or therapeutic treatment of Chlamydial

CC bacterial infection (specifically those caused by C. pneumoniae,

CC C. psittaci or C. trachomatis), e.g. sinusitis, pharyngitis, bronchitis,

CC pneumonia, asthmatic bronchitis, adult-onset asthma, chronic

CC obstructive pulmonary disease (COPD), atherosclerosis or

CC The sequences shown in ABP61993-ABP62011 represent proteins BVH-CPN1 to

CC BVH-CPN18 of C. pneumoniae.

XX Sequence 310 AA;

Query Match 40.9%; Score 663; DB 23; Length 310;

Best Local Similarity 45.0%; Pred. No. 9.9e-53;  
Matches 143; Conservative 59; Mismatches 98; Indels 18; Gaps 3;

CC 2 QIVLAMPKPGCAGVDRALEIVDOALTEAFGAPVYRHEVYHNRVTVDGKQKGAFTIEELS 61

DB 3 KLILCNPRGPGSGVVRALQVVEVLEKMGAPVYVKEHIVNHRVYVVALAKAGALFVEELV 62

QY 62 DVPVGSYLIFSAHAGSKKEVQGEAEEROLTVFADTCPLVTXHMVAKHAKQGREVILIGH 121

DB 63 DVEGERVIYSAGHGPSVRARAKRLIDIDATCGLVTKHSAKLYASRGYKILILIGH 122

QY 122 AGHPEVEGTMQYKTEGGGIYLVETPEDEVRLKVNPNDAVYVQTLLSMTDTKAWD 181

DB 123 KKHVEVIGIVEVPE-----HITVEKVADEVALPFSSDPLFYITQTLLSDVQGEISS 177

QY 182 ALREQPSIIEQKDDICATQNRQDAVHDLAKISDLILVYSGPSNSNRLREIAYQLG 241

DB 178 ALTKRPSIITLPSSICATITNRQKALRSVLSRVNYVYVGVNNSNRLREVALRGG 237

QY 242 KPAVLIIDTYODLKQDMLGIEVGVGTAGASAPVAVQEVLDQKAMGETTSVRANSGLIE 301

DB 238 VPADLIINNPEDIDITNVHSGDIAMTAGASTPEDVVOACIRKL-----SSLIGLQVE 290

QY 302 -----EKVYFSPKELK 313

DB 291 NDIFAVEDVFOLPKELR 308

RESULT 13

AAU09434

ID AAU09434 standard; Protein; 310 AA.

AC AAU09434;

DT 26-MAR-2002 (first entry)

XX Chlamydia pneumoniae metalloproteinase.

XX ATP binding cassette; secretory locus open reading frame; endopeptidase;

XX secretory locus ORF; protease; metalloproteinase; CLP protease Atrase;

XX CLP protease subunit; transglycolase/transpeptidase; CLPc protease;

XX chirodoxin; Chlamydia infection; antibacterial.

XX Chlamydia pneumoniae CWL029.

XX WO200185972-A2.

XX 15-NOV-2001.

XX 08-MAY-2001; 2001WO-CA00653.

XX 08-MAY-2000; 2000US-202672P.

XX 30-MAY-2000; 2000US-207852P.

XX 16-JUN-2000; 2000US-211796P.

XX 16-JUN-2000; 2000US-211797P.

XX 16-JUN-2000; 2000US-211798P.

XX 16-JUN-2000; 2000US-211801P.

XX 16-JUN-2000; 2000US-212044P.

CC The present invention relates to the isolation of Chlamydia  
CC pneumoniae strain CW029 genes and their encoded proteins. The genes of  
CC the invention encode an ATP binding cassette gene, a secretory locus  
CC open reading frame (ORF), an endopeptidase, a protease, a  
CC metalloprotease, a Clp protease ATPase, a Clp protease subunit,  
CC transglycolase/transpeptidase, a ClpC protease, or thioresdoxin. The  
CC genes of the invention can be used in a vector as a vaccine for the  
CC prevention and treatment of Chlamydia infections. Also described are  
CC B- and T-cell epitopes from the proteins of the invention which can  
CC be used as Chlamydia antigens. AAU0430-AAU09439 represent the proteins  
CC encoded by the C. pneumoniae genes (AA518750-AA518755) of the invention  
XX  
XX Sequence 310 AA;  
XQ

RESULT 14	
ID	ABP66088
XX	ABP66088 standard; Protein; 352 AA.
XX	
AC	ABP66088;
XX	
DT	19-NOV-2002 (first entry)
XX	
DE	Bifidobacterium longum NCC2705 ORF amino acid sequence SEQ ID NO:832.
XX	
XX	
KM	Bifidobacterium longum NCC2705; Bifidobacterium; bacterial;
KW	antidiarrheic; antibacterial; inhibitor of Salmonella; detection;
KW	identification; lactic acid bacterium; diarrhoea; pathogenic bacteria

CC The present invention describes a polynucleotide (I) comprising a  
CC sequence of a Bifidobacterium genome selected from the nucleotide  
CC sequences given in AB081842 and AB081943, or a sequence exhibiting at  
CC least 90% identity or which hybridises with the sequences given in  
CC AB081942 and AB081843. Also described is a polynucleotide (II) encoding  
CC a fusion protein, comprising a sequence selected from 1097 sequences  
CC given in AB065558 to AB065154, ligated in frame to a polynucleotide  
CC encoding a heterologous polypeptide. (I) has antidiarrhetic and  
CC antibacterial activities, and can be used as an inhibitor of *Salmonella*  
CC (I) (which is a probe) is useful for the detection and/or identification  
CC of Bifidobacterium longum in a biological sample. A carrier containing  
CC the lactic acid bacterium Bifidobacterium longum NCC2705 (CNCM 1-2618)  
CC can be used for preventing and/or treating diarrhoea brought about by  
CC pathogenic bacteria and/or rotavirus. The carrier is a food composition  
CC selected from milk, yogurt, curd, cheese, fermented milks, milk based  
CC fermented products, ice-creams, fermented cereal based products, milk  
CC based powders, infant formula, pet food or a pharmaceutical composition  
CC selected from tablets, liquid bacterial suspensions, dried oral  
CC supplement, wet oral supplement, dry tube feeding or wet tube feeding.  
CC (I) is useful in DNA arrays or chips to carry out analysis of the  
CC expression of the Bifidobacterium gene. AB081844 to AB081850 represent  
CC Bifidobacterium related nucleotide sequences given in the Sequence  
CC listing from the present invention but not mentioned further within the  
CC specification.  
CC N.B. The sequence data for this patent is not represented in the printed  
CC specification but is based on sequence information supplied by the  
CC European Patent Office.

Query Match	40.5%	Score 658	DB 23	Length 352
Best Local Similarity	43.8%	Pred. No. 3,56-52		
Matches 145	Conservative 64	Mismatches 96	Indels 26	Gaps 7
QY	3	YLAPRFPCGVRAI---	EIVDQALEAG-----	APIYRELVNHRVTVDGLKQ 51
		: : : : : :	: : : : : :	: : : : : :
Db	27	VVLADPRFCAGVDRAILVTQTILKAELAAGKRTBDGLPRVYVRQIYHNKKNVEDLNG 86		
QY	52	KAVVIELLSVLP-----	VGSYLLFSKNGSKSKEVQQAERQLTVDPALCPYLTAKHMY 106	
		: : : : : :	: : : : : :	: : : : : :
Db	87	QCAVNVQQLAEIPALAAQAGIPVFSFASHGSPVYKKAERGRHGVADATCPYKXKHEHV 146		
QY	107	AKHAGKEFVLLIGHAGPEVEGTMGQYEKCTGGGIVLVEIPEDVRYNKKVNPDLATV 166		
		: : : : : :	: : : : : :	: : : : : :
Db	147	LFVREGEYIYIGKHGDVAIVGGSPE-----	VHYHLIHEHSDVSDLPAPDPTLYLL 201	
QY	167	TQTLSTMTDKRMVDAREQRPSPKQKEDICLAQNQODVAHDLAKLSDLILVYVGSN 226		
		: : : : : :	: : : : : :	: : : : : :
Db	202	SOITLSVDEIMDTTAAKKAFPMWIOEPSSDICYLAISNQAQKVLAAQSDCCVIVGSAN 261		
QY	227	SNSNRRLREIAVQ--LGR--	PAYLIDTVQDKQDMLEGIIRVGVGTAGASAPVLYVQDEVIDQ 283	

DB 262 SSSNRLKLVKAGGCGEERKAYRVDASGLDPAWFEGLSGLSGASVDELVSGLVDA 321  
 QY 284 LKAWGETTSVRENGIEERKVFVSIPKELKK 314  
 DB 322 LQNLG--FTGMKSVETIKENMFVLPALRR 350

RESULT 15  
 AAY37259

ID AAY37259 standard; Protein; 307 AA.

AC AAY37259;

DT 07-OCT-1999 (first entry)

DE Amino acid sequence of a Chlamydia trachomatis protein.

XX Vaccine; eye disease; conventional trachoma; nonendemic trachoma;

KM paratrachoma; inclusion conjunctivitis; genital disease; perihepatitis;

KM nongonococcal urethritis; epididymitis; cervicitis; salpingitis;

KM bartolinitis; pneumonia; venereal lymphogranulomatosis.

XX Chlamydia trachomatis.

OS WO928475-A2.

XX 10-JUN-1999.

PD 27-NOV-1998; 98MO-IB01939.

PR 04-NOV-1998; 98US-0107077.

PR 28-NOV-1997; 97FR-0015041.

PR 17-DEC-1997; 97FR-0016034.

XX (GBST) GENSET.

PI Griffiths R;

DR WPI; 1999-371125/31.

PT Genome sequence of Chlamydia trachomatis

PS Disclosure; Page 1012-1013; 1755pp; English.

XX AAY36754-Y37949 are encoded by open reading frames (ORFs) of the genome

CC of Chlamydia trachomatis (see AAY01425). The polypeptides can be used as

CC vaccines against Chlamydia trachomatis. Antisense and ribozyme sequences

CC can also be used to control growth of the microorganism. Chlamydia

CC trachomatis is responsible for a large number of diseases, e.g. eye

CC diseases such as conventional trachoma, nonendemic trachoma,

CC paratrachoma, and inclusion conjunctivitis; genital diseases such as

CC nongonococcal urethritis, epididymitis, cervicitis, salpingitis,

CC perihepatitis, bartolinitis; pneumonia; venereal lymphogranulomatosis;

CC and venereal lymphogranulomatosis. The polypeptides of the invention

CC may be of use in treating these diseases.

XX SQ Sequence 307 AA;

QY 2 QIVLANPFGCGVNDRAIEYDQAIETAGAPITYREHVVNRTTYDGLKQKAGVIEELS 61

DB 3 KILISPRFGAGVIRAIQIVVALEKMRPIYVHREIVHNNHVDKLRKGAIRIEDLQ 62

QY 62 DVPVSSYLFFSAHGSKEVQEAERQUTLVFATCPLVTKVMQVAKAKQGREVILIGH 121

DB 63 EVPRRSRVIFSHGVPSPAREATERGLAIDATGVLTKHSAVKMRAKKGYHILLIGK 122

QY 122 AGHPEVEGTMGQIEKCTBSGGIYLVETPEDVFNKVNNDLAVYTQTLTNTDTKAWVD 181

DB 123 RKHVEILIGRGE-----APDQITVENIAEVEALPESAQDPLFYVYQTLTSMDDAADIYA 177  
 QY 182 ALRQFSPSIRKQKDDI CYATQNRQDAVHDLAKISDLILVVGSSNSNRLAEIAYQLG 241  
 DB 178 ALKARVPRIFTLPSSSICATQNRQDALRNLIPQVDFTVYVIGDRQSSNSNRLAEIAYRERG 237  
 QY 242 KPAVLIDITYQDLKQDNLGIEVGVGTAGASAPETLY-----QEVLDQKAWGETTSV 294  
 DB 238 VTARLVNHPDEVTEILLQYSGNIGITAGASTPBDVQACIMKQELLIPDL-----SI 289  
 QY 295 RENGIEERKVFVSIPKEL 312  
 DB 290 EMDLFEVBDITVFQLPKEL 307

Search completed: January 29, 2004, 15:49:46  
 Job time : 30.0679 secs

GenCore version 5.1.6  
Copyright (c) 1993 - 2004 CompuGen Ltd.

OW protein - protein search, using sw model

Run on: January 29, 2004, 15:45:34 : Search time 9.98038 Seconds  
(without alignments)  
1348.130 Million cell updates/sec

Title: US-09-941-947a-18  
Perfect score: 1623  
Sequence: 1 MQIVLANPRGFCAGVDRAIE.....GIEKVFSPKELKGMQA 318

Scoring table: BLOSUM62  
Gapop 10.0 , Gapext 0.5

Searched: 328717 seqs, 42310858 residues

Total number of hits satisfying chosen parameters: 328717

Minimum DB seq length: 0  
Maximum DB seq length: 200000000

Post-Processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Database : Issued Patents AA:  
1: /cgn2\_6/prodata/1/iaa/5A\_COMB.pep:\*  
2: /cgn2\_6/prodata/1/iaa/5B\_COMB.pep:\*  
3: /cgn2\_6/prodata/1/iaa/6A\_COMB.pep:\*  
4: /cgn2\_6/prodata/1/iaa/6B\_COMB.pep:\*  
5: /cgn2\_6/prodata/1/iaa/6C\_COMB.pep:\*  
6: /cgn2\_6/prodata/1/iaa/backfill61.pep:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	1048.5	64.6	336	4	US-09-328-352-6046 Sequence 6046, Ap
2	1028.5	63.4	349	4	US-09-199-6378-429 Sequence 429, App
3	1028.5	63.4	368	4	US-09-252-991A-18616 Sequence 24, Appl
4	663	40.9	310	4	US-09-198-452A-1090 Sequence 1090, Ap
5	116.5	7.2	521	4	US-09-107-532A-6431 Sequence 6431, Ap
6	104	6.4	494	4	US-09-134-001C-5024 Sequence 5024, Ap
7	103	6.3	586	4	US-08-198-446B-19 Sequence 19, Appl
8	103	6.3	586	4	US-08-870-693-19 Sequence 19, Appl
9	101.5	6.3	540	4	US-09-134-001C-4705 Sequence 4705, Ap
10	101	6.2	547	2	US-08-467-822-35 Sequence 35, Appl
11	101	6.2	547	3	US-08-432-697-35 Sequence 35, Appl
12	101	6.2	547	3	US-08-466-248-35 Sequence 35, Appl
13	99.5	6.1	1196	2	US-08-144-121-4 Sequence 4, Appl
14	99.5	6.1	1196	2	US-08-735-893-4 Sequence 4, Appl
15	95.5	5.9	643	3	US-08-797-358B-3 Sequence 3, Appl
16	95	5.9	730	3	US-09-398-865A-2 Sequence 2, Appl
17	95	5.8	730	4	US-09-710-714-2 Sequence 2, Appl
18	94.5	5.8	516	4	US-09-134-001C-3404 Sequence 3404, Ap
19	94	5.8	496	4	US-09-202-491-7 Sequence 7, Appl
20	93	5.7	573	2	US-08-706-209-1 Sequence 1, Appl
21	93	5.7	573	3	US-08-981-787-1 Sequence 1, Appl
22	93	5.7	573	4	US-08-461-722-1 Sequence 1, Appl
23	93	5.7	573	4	US-08-336-251-1 Sequence 1, Appl
24	93	5.7	573	4	US-09-468-041-1 Sequence 1, Appl
25	93	5.7	573	4	US-08-381-861-1 Sequence 1, Appl
26	93	5.7	573	4	PCT-US94-06362-1 Sequence 1, Appl
27	93	5.7	573	5	PCT-US96-11373-1 Sequence 1, Appl

28	93	5.7	573	5	PCT-US96-11373-1 Sequence 1, Appl
29	93	5.7	981	4	US-09-252-991A-18616 Sequence 18616, A
30	91.5	5.6	298	3	US-08-961-083-24 Sequence 24, Appl
31	91.5	5.6	298	4	US-09-536-784-24 Sequence 24, Appl
32	91.5	5.6	2548	4	US-09-172-442-1 Sequence 1, Appl
33	91	5.6	335	4	US-09-934-901-6 Sequence 6, Appl
34	90.5	5.6	540	4	US-08-461-722-3 Sequence 3, Appl
35	90.5	5.6	540	4	US-08-336-251-3 Sequence 3, Appl
36	90.5	5.6	540	4	US-09-468-041-3 Sequence 3, Appl
37	90.5	5.6	540	5	PCT-US94-06362-3 Sequence 3, Appl
38	90.5	5.6	541	2	US-08-467-822-34 Sequence 34, Appl
39	90.5	5.6	541	2	US-08-447-154-19 Sequence 19, Appl
40	90.5	5.6	541	3	US-08-432-697-34 Sequence 34, Appl
41	90.5	5.6	541	3	US-08-466-248-34 Sequence 34, Appl
42	90.5	5.6	710	4	US-09-107-532A-5067 Sequence 5067, Ap
43	89.5	5.5	243	4	US-09-252-991A-18154 Sequence 18154, A
44	88.5	5.5	885	2	US-08-310-912A-2 Sequence 2, Appl
45	88.5	5.5	885	3	US-08-841-089-2 Sequence 2, Appl

ALIGNMENTS

RESULT 1  
US-09-328-352-6046  
; Sequence 6046, Application US/09328352  
; Patent No. 6562958  
; GENERAL INFORMATION:  
; APPLICANT: Gary L. Breton et al.  
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO ACINETOBACTER  
; FILE REFERENCE: GTC99-03PA  
; CURRENT APPLICATION NUMBER: US/09/328,352  
; CURRENT FILING DATE: 1999-06-04  
; NUMBER OF SEQ ID NOS: 8252  
; SEQ ID NO 6046  
; TYPE: PRT  
; ORGANISM: Acinetobacter baumannii  
US-09-328-352-6046

Query Match 64.6%; Score 1048.5; DB 4; Length 336;  
Best Local Similarity 64.2%; Pred. No. 2.5e-98;  
Matches 203; Conservative 49; Mismatches 61; Indels 3; Gaps 2;  
QY 1 MQIVLANPRGFCAGVDRAIEIVDDAIEAFGAPIYVREHVANRTVVDGKQGAFFIEL 60  
DB 21 MEIVLANPRGFCAGVDRAIIVNRALECFNPPIYRREHVANKFFVDDLRQGAFFVEL 80  
QY 61 SDVPVGSYLIFSAGVSKVQOQAEEROLTYFDATCPVTVKQVAVAHAKQREVLIG 120  
DB 81 DQVPDSDIVISAGVSAVQOQAEEROLTYFDATCPVTVKQVAVAHAKQREVLIG 140  
QY 121 HAGHEVGVGNGQVEKCTGIVLVEPTEDVNRKVNPNDLAVYTQTLSTMTDKVMV 180  
DB 141 HEGHPEVGVGNGQVQDK-LKGGDIYHVEDEADVALLVGRPEKLAFTYQTLSTIDTAIVY 199  
QY 161 DALRQPSIYKQKKDDICVATQNRQDAVHDLAKISDLILVGSFNSNSRLREIAVOL 240  
DB 200 DALRQPSIYKQKKDDICVATQNRQDAVHDLAKISDLILVGSFNSNSRLREIAVOL 259  
QY 241 GKPAVLITQYDQKQDMIEGVGVTAQAPVLAQVLEVDQKAWGSETTSVSENGSI 300  
DB 260 GKPAVLITQYDQKQDMIEGVGVTAQAPVLAQVLEVDQKAWGSETTSVSENGSI 317  
QY 301 EEKVFSPKELKGM 316  
DB 318 EENIFSPKELRIHV 333  
RESULT 2  
US-09-199-637A-429  
; Sequence 429, Application US/09199637A

```

; Patent No. 6355411
; GENERAL INFORMATION:
; APPLICANT: Ausubel, Frederick
; APPLICANT: Goodman, Howard W.
; APPLICANT: Rahme, Laurence G.
; APPLICANT: Mahajan-Miklos, Shalina
; APPLICANT: Tan, Man-Wah
; APPLICANT: Cao, Hui
; APPLICANT: Drenkard, Eliana
; APPLICANT: Tsongalis, John
; TITLE OF INVENTION: VIRULENCE-ASSOCIATED NUCLEIC ACID
; FILE REFERENCE: 00786/361002
; CURRENT APPLICATION NUMBER: US/09/199,637A
; PRIOR FILING DATE: 1998-11-25
; PRIOR APPLICATION NUMBER: 60/066,517
; PRIOR FILING DATE: 1997-11-25
; NUMBER OF SEQ ID NOS: 437
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 429
; LENGTH: 349
; TYPE: PR1
; ORGANISM: Pseudomonas aeruginosa
US-09-199-637A-429

```

```

Query Match 63.4%; Score 1028.5; DB 4; Length 349;
Best Local Similarity 63.3%; Pred. No. 2,9e-96;
Matches 198; Conservative 49; Mismatches 63; Indels 3; Gaps 2;

```

```

QY 1 MOVLANPRGFCAGVDRALIEIVDOAIEAFGAPTYRHHVHNRTVVDGLKOGKAVFIEEL 60
DB 36 MOVLANPRGFCAGVDRALIEIVDOAIEAFGAPTYRHHVHNRTVVDGLKOGKAVFIEEL 95
QY 61 SDVPVGSYLIFSAHGVSKVQDAEROLTVDPATCPITYKVMQVAKHAKQGREVILIG 120
DB 96 DVPDNVIVIFSAHGVSKVQDAEROLTVDPATCPITYKVMQVAKHAKQGREVILIG 155
QY 121 HAHPEVEGTMGQYKCTEGGGIYIVETPEDVBNLKNPNNDLAVYTQTTLSTMTKXWV 180
DB 156 HAHPEVEGTMGQYD-ASNGGAIYVDEADVAALEVRKPEALHYTQTTLSTMTSKYI 214
QY 181 DALREQPSIKEOKKDDICATONRODAVHDLAKISDLILVGSPPSSNSNRLREIAVOL 240
DB 215 DALRAKFPQIGPRKNDICATONRODAVHDLAKISDLILVGSPPSSNSNRLREIAVOL 274
QY 241 GKPAVILDTYQDLKQDWLEIGVGVTVAGASAPVTVQEVTDOLKAMGSETTSVRENSGI 300
DB 275 GTPAYLIDGABDMQGMDFGVRIIGITAGASAPVTVQEVTVQEVTDOLKAMGSETTSVRENSGI 332
QY 301 EEKVFESIPKEIK 313
DB 333 EENITFSMKELR 345

```

## RESULT 3

```

US-09-252-991A-21399
; Sequence 21399, Application US/09252991A
; Patent No. 6551795

```

```

; GENERAL INFORMATION:
; APPLICANT: Marc J. Rubenfield et al.
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO PSEUDOMONAS
; FILE REFERENCE: 107196, 136
; CURRENT APPLICATION NUMBER: US/09/252,991A
; PRIOR FILING DATE: 1999-02-18
; PRIOR APPLICATION NUMBER: US 60/074,788
; PRIOR FILING DATE: 1998-02-18
; PRIOR APPLICATION NUMBER: US 60/094,190
; NUMBER OF SEQ ID NOS: 33142
; SEQ ID NO 21399
; LENGTH: 368
; TYPE: PR1

```

```

; ORGANISM: Pseudomonas aeruginosa
US-09-252-991A-21399

```

```

Query Match 63.4%; Score 1028.5; DB 4; Length 368;
Best Local Similarity 63.3%; Pred. No. 3.2e-96;
Matches 198; Conservative 49; Mismatches 63; Indels 3; Gaps 2;

```

```

QY 1 MOVLANPRGFCAGVDRALIEIVDOAIEAFGAPTYRHHVHNRTVVDGLKOGKAVFIEEL 60
DB 55 MOVLANPRGFCAGVDRALIEIVDOAIEAFGAPTYRHHVHNRTVVDGLKOGKAVFIEEL 114
QY 61 SDVPVGSYLIFSAHGVSKVQDAEROLTVDPATCPITYKVMQVAKHAKQGREVILIG 120
DB 115 DVPDNVIVIFSAHGVSKVQDAEROLTVDPATCPITYKVMQVAKHAKQGREVILIG 174
QY 121 HAHPEVEGTMGQYKCTEGGGIYIVETPEDVBNLKNPNNDLAVYTQTTLSTMTKXWV 180
DB 175 HAHPEVEGTMGQYD-ASNGGAIYVDEADVAALEVRKPEALHYTQTTLSTMTSKYI 233
QY 181 DALREQPSIKEOKKDDICATONRODAVHDLAKISDLILVGSPPSSNSNRLREIAVOL 240
DB 234 DALRAKFPQIGPRKNDICATONRODAVHDLAKISDLILVGSPPSSNSNRLREIAVOL 293
QY 241 GKPAVILDTYQDLKQDWLEIGVGVTVAGASAPVTVQEVTDOLKAMGSETTSVRENSGI 300
DB 294 GTPAYLIDGABDMQGMDFGVRIIGITAGASAPVTVQEVTVQEVTDOLKAMGSETTSVRENSGI 351
QY 301 EEKVFESIPKEIK 313
DB 352 EENITFSMKELR 364

```

## RESULT 4

```

US-09-198-452A-1090
; Sequence 1090, Application US/09198452A
; Patent No. 6559284

```

```

; GENERAL INFORMATION:
; APPLICANT: Griffiths, R.
; TITLE OF INVENTION: Chlamydia pneumoniae genomic sequence and polypeptides, fragment;
; TITLE OF INVENTION: thereof and uses thereof, in particular for the diagnosis, prev
; FILE REFERENCE: 9710-003-999
; CURRENT APPLICATION NUMBER: US/09/198,452A
; PRIOR FILING DATE: 1998-11-24
; NUMBER OF SEQ ID NOS: 6849
; SEQ ID NO 1090
; LENGTH: 310
; TYPE: PR1
; ORGANISM: Chlamydia pneumoniae
US-09-198-452A-1090

```

```

Query Match 40.9%; Score 663; DB 4; Length 310;
Best Local Similarity 45.0%; Pred. No. 3.7e-59;
Matches 143; Conservative 59; Mismatches 98; Indels 18; Gaps 3;

```

```

QY 2 OIVLANPRGFCAGVDRALIEIVDOAIEAFGAPTYRHHVHNRTVVDGLKOGKAVFIEEL 61
DB 3 KLILNPRGFCAGVDRALIEIVDOAIEAFGAPTYRHHVHNRTVVDGLKOGKAVFIEEL 62
QY 62 DVPVGSYLIFSAHGVSKVQDAEROLTVDPATCPITYKVMQVAKHAKQGREVILIG 121
DB 63 DVPESERVIVSAHGIPEVRAKAKRIDIDATCGVTVKHSAAKYASGGYKILIGH 122
QY 122 AGHPEVEGTMGQYKCTEGGGIYIVETPEDVBNLKNPNNDLAVYTQTTLSTMTKXWV 181
DB 123 KKEVAVIGIVGEVBE-----HITVEKADVBEALPSSDTPLFTITQTTLSDVOEISS 177
QY 182 ALREQPSIKEOKKDDICATONRODAVHDLAKISDLILVGSPPSSNSNRLREIAVOLG 241
DB 178 ALKRVPSIITLPSSTICATYATNROKALRSYLSRVYVYVGVGVNSNSNRLREIALRG 237
QY 242 KPAVILDTYQDLKQDWLEIGVGVTVAGASAPVTVQEVTDOLKAMGSETTSVRENSGI 301
DB 242 KPAVILDTYQDLKQDWLEIGVGVTVAGASAPVTVQEVTDOLKAMGSETTSVRENSGI 301

```

Db 238 VPADLNNPEDIIVNHSGLIANTAGASTPEVYQACIRL-----SSLIPGLQVE 290

QY 302 -----EYVFSIRPELK 313

Db 291 NDIPAEVDVFPOLPELR 308

RESULT 5

US-09-107-532A-6431

Sequence 6431, Application US/09107532A

Patent No. 6583275

GENERAL INFORMATION:

APPLICANT: Lynn A Doucette-Stamm and David Bush

TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO ENTEROCOCCUS FAECIUM FOR DIAGNOSTICS AND THERAPEUTICS

NUMBER OF SEQUENCES: 7310

CORRESPONDENCE ADDRESS:

ADDRESSEE: GENOME THERAPEUTICS CORPORATION

STREET: 100 Beaver Street

CITY: Waltham

STATE: Massachusetts

COUNTRY: USA

ZIP: 02354

COMPUTER READABLE FORM:

MEDIUM TYPE: CD-ROM ISO9660

COMPUTER: PC

OPERATING SYSTEM: <Unknown>

SOFTWARE: ASCII

CURRENT APPLICATION DATA:

APPLICATION NUMBER: US/09/107,532A

FILING DATE: 30-Jun-1998

PRIOR APPLICATION DATA:

APPLICATION NUMBER: 60/085,598

FILING DATE: 14 May 1998

APPLICATION NUMBER: 60/051571

FILING DATE: July 2, 1997

ATTORNEY/AGENT INFORMATION:

NAME: Arinello, Pamela Deneke

REGISTRATION NUMBER: 40,489

REFERENCE/DOCKET NUMBER: GTC-012

TELECOMMUNICATION INFORMATION:

TELEPHONE: (781)893-5007

TELEFAX: (781)893-8277

INFORMATION FOR SEQ ID NO: 6431:

SEQUENCE CHARACTERISTICS:

LENGTH: 521 amino acids

TYPE: amino acid

TOPOLOGY: linear

MOLECULE TYPE: protein

HYPOTHEICAL: YES

ORIGINAL SOURCE:

ORGANISM: Enterococcus faecium

FEATURE:

NAME/KEY: misc feature

LOCATION: (B) LOCATION 1...521

SEQUENCE DESCRIPTION: SEQ ID NO: 6431:

US-09-107-532A-6431

Query Match

Best Local Similarity 7.2%; Score 116.5; DB 4; Length 521;

Matches 63; Conservative 52; Mismatches 104; Indels 83; Gaps 11;

QY 38 EYVNRRTVYDGLKQAGVFIELSDVPGSY-----LIFSAGVSKVQQAEEQQLT 90

Db 112 EVVHTSLVBNKIKISFIAGLQPSNDLVYDLNLIPTKHNSRNLIQTTRK---- 167

QY 91 VFDAICPLVTVHNVAKQAGREVILIGHAGPEVEG--TWQYKCTEGGSIYVER 148

Db 168 -----VPI-----VILEGKTGLSIQPIYSKTRKIGVAGSPYLLSS 206

QY 149 PEDVRLKVNPNDLAVYTQTLSKTDTKW-----VDALRQPPSIRKQKDD 197

Db 207 FYDIRNKL-----LTIIVLEVSVLLSIISIGFLSSYFLKPLKVLADTDITRKQPSD 261

QY 198 ICYATONRODAVDIACISLILVSGPNSNSNRRL-----BIAVOLGKAYL 246

Db 262 IMHEIDTNDLADLAFENEL-----DRMLYLIEQGEVQFVDSHELRTPVAI 311

QY 247 IDTYQDLKQDM-LEGIEVGVTAGASAPV-----LVQEVTDOLKA-----MGQETTSV 294

Db 312 IEGLSLINRNGDDPILBESLSASQWISRKSLVQENLDISRAQVYVHVRNTTKG 371

QY 295 RE 296

Db 372 KE 373

RESULT 6

US-09-134-001C-5024

Sequence 5024, Application US/09134001C

Patent No. 6380370

GENERAL INFORMATION:

APPLICANT: Lynn Doucette-Stamm et al

TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO STAPHYLOCOCCUS

FILE REFERENCES: GTC-007

CURRENT FILING DATE: 1998-08-13

PRIOR APPLICATION NUMBER: US 60/064,964

PRIOR FILING DATE: 1997-11-08

PRIOR APPLICATION NUMBER: US 60/055,779

PRIOR FILING DATE: 1997-08-14

NUMBER OF SEQ ID NOS: 5674

SEQ ID NO 5024

LENGTH: 494

TYPE: PRT

ORGANISM: Staphylococcus epidermidis

US-09-134-001C-5024

Query Match

Best Local Similarity 6.4%; Score 104; DB 4; Length 494;

Matches 64; Conservative 63; Mismatches 134; Indels 74; Gaps 12;

QY 52 KGAVFIEELSDVPGSYLIFSAGVSKVQQAEEQQLTFFDQTCPLTVKHVQVAGAK 111

Db 14 KESLTFDDVLLIPASDVLPSSVDVLSVKLSDKI-KNIPVTSAGMDTVTSKRALAK-AR 71

QY 112 QGSEVILIGHAG-----HP-----EYEGTWQYKCTEGG 141

Db 72 QGGLGVYHKKMGVEEQADDEVQYKRSNGVYSNPFLLTPRESVYBAALMGKTRI-----S 127

QY 142 GYLVATPEPVRKLVKNNPNDLAVYTQTLSMTD-----TKWVDALRQPPSIRKQ 193

Db 128 GVPIDVNOEDRKILGILTNRDLRFIEDFSIKISDVTKDNLTAPVGTTLDEBALIQKH 187

QY 194 KQDICIYATONROD--AVHDLAKISDL-----ILVSGPNSNSNRRLR-BIAVQ 239

Db 188 KIRKLPIVNRKRLGSLITKDIKYLEFPYAADHGRLLAAIGSDVTSKOTELRACQVLE 247

QY 240 LGRPAVLIPT-----YQDLKQDMLEGIEVGVTAGASAPVYLVOEVTIDOLKAWG 288

Db 248 AGVDALIIDTAHSGSKGVINQVGHIKETYPETIVVAGNVATAETALPEAGADVAVVGI 307

QY 289 GE-----TTSVENSIGEE-KVPSIRPELKQMOA 318

Db 308 GPGSICITRRVAGVGPQITAVYDCAETEARKGKA 342

RESULT 7

US-08-198-446B-19

Sequence 19, Application US/08198446B

Patent No. 5674966

GENERAL INFORMATION:

APPLICANT: Hartwell, Leland B.

APPLICANT: Weinstein, Ted A.

APPLICANT: Pilon, Sharon E.

APPLICANT: Groudine, Mark T.  
TITLE OF INVENTION: Cell Cycle Checkpoint Genes  
NUMBER OF SEQUENCES: 19  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: Christensen O'Connor Johnson Kindness PLLC  
STREET: 1420 Fifth Ave., Suite 2800  
CITY: Seattle  
STATE: WA  
COUNTRY: USA  
ZIP: 98101-2347  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: PatentIn Release #1.0, Version #1.25  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/198,446B  
FILING DATE: 18-Feb-1994  
CLASSIFICATION: 536  
ATTORNEY/AGENT INFORMATION:  
NAME: Sheiness, Diana K.  
REGISTRATION NUMBER: 35,356  
REFERENCE/DOCKET NUMBER: FHCRI17537  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: 206-682-8100  
TELEFAX: 206-224-0779  
INFORMATION FOR SEQ ID NO: 19:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 586 amino acids  
TYPE: amino acid  
TOPOLOGY: linear  
MOLECULE TYPE: protein  
DESCRIPTION: Yeast MEC3 protein  
US-08-198-446B-19

Query Match 6.3%; Score 103; DB 1; Length 586;  
Best Local Similarity 21.9%; Pred. No. 0.1;  
Matches 71; Conservative 46; Mismatches 97; Indels 110; Gaps 18;

34 YVRHE--VVRNRVTVDGLKQKGAVFIEELSDVPVGSYLIFSA-----HGVS-----77  
15 YRHAFLRYNHTTPAQKQA-----QIQIPIENYRNFSTVAHVHGKSTLSRLLEI 69

78 -----KEVQGEAE---ERQLTVFDPATCPPLVTKVMQVAKAKQREVI--LIQHA 122  
70 TVHIDPNARKQVLDKLEVERERGITTKAQTCM-----FYKDKRTGKRYLHLIDTP 122

123 GHPEVEGTMGQ--YEKCTEGGIGIVLVEPDEVRLKVNPNNDLAY-----VTQTLLS 172  
123 GHVDFRGVSRSYASC--GGAILLVDSAGCIQAQTVAN--FYLFSLGLKLPVINKIDLN 179

173 MDTKVMVDALRROPFSIKQKDDICATQNRQDAVHDL--AKISDLILVVGSSPNSSN 229  
180 FTDVQKQKQIYVNNF-----ELPREDIIGVSRKTALNVEELLPAIIDRIPEPTGRPD--- 232

230 SNRLREIAVOLGRP--AYLIDTYQDLKQDWLBGIEVVGTVAGASAPVLYQGEVTDQKAM 287  
233 -----KPRRLVDSWYD-----AYLGAVLVNIVDGF--- 260

288 GGETSVRENSGIEBKVPSIPKE 311  
261 -----VRKN-----DKVICAQTKK 274

RESULT 8  
US-08-870-693-19  
Sequence 19, Application US/08870693  
GENERAL INFORMATION:  
APPLICANT: Bartwell, Ieland H.  
APPLICANT: Weinert, Ted A.  
APPLICANT: Pion, Sharon E.  
APPLICANT: Groudine, Mark T.

TITLE OF INVENTION: Cell Cycle Checkpoint Genes  
NUMBER OF SEQUENCES: 19  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: Christensen O'Connor Johnson Kindness PLLC  
STREET: 1420 Fifth Ave., Suite 2800  
CITY: Seattle  
STATE: WA  
COUNTRY: USA  
ZIP: 98101-2347  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: PatentIn Release #1.0, Version #1.25  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/870,693  
FILING DATE:  
CLASSIFICATION: 536  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US 08/198,446  
FILING DATE: February 18, 1994  
APPLICATION NUMBER: PCT/US93/04458  
FILING DATE: May 12, 1993  
APPLICATION NUMBER: US 07/884,426  
FILING DATE: May 14, 1992  
APPLICATION NUMBER: US 07/882,051  
FILING DATE: May 12, 1992  
ATTORNEY/AGENT INFORMATION:  
NAME: Sheiness, Diana K.  
REGISTRATION NUMBER: 35,356  
REFERENCE/DOCKET NUMBER: FHCRI10798  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: 206-682-8100  
TELEFAX: 206-224-0779  
INFORMATION FOR SEQ ID NO: 19:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 586 amino acids  
TYPE: amino acid  
TOPOLOGY: linear  
MOLECULE TYPE: protein  
DESCRIPTION: Yeast MEC3 protein  
US-08-870-693-19

Query Match 6.3%; Score 103; DB 2; Length 586;  
Best Local Similarity 21.9%; Pred. No. 0.1;  
Matches 71; Conservative 46; Mismatches 97; Indels 110; Gaps 18;

34 YVRHE--VVRNRVTVDGLKQKGAVFIEELSDVPVGSYLIFSA-----HGVS-----77  
15 YRHAFLRYNHTTPAQKQA-----QIQIPIENYRNFSTVAHVHGKSTLSRLLEI 69

78 -----KEVQGEAE---ERQLTVFDPATCPPLVTKVMQVAKAKQREVI--LIQHA 122  
70 TVHIDPNARKQVLDKLEVERERGITTKAQTCM-----FYKDKRTGKRYLHLIDTP 122

123 GHPEVEGTMGQ--YEKCTEGGIGIVLVEPDEVRLKVNPNNDLAY-----VTQTLLS 172  
123 GHVDFRGVSRSYASC--GGAILLVDSAGCIQAQTVAN--FYLFSLGLKLPVINKIDLN 179

173 MDTKVMVDALRROPFSIKQKDDICATQNRQDAVHDL--AKISDLILVVGSSPNSSN 229  
180 FTDVQKQKQIYVNNF-----ELPREDIIGVSRKTALNVEELLPAIIDRIPEPTGRPD--- 232

230 SNRLREIAVOLGRP--AYLIDTYQDLKQDWLBGIEVVGTVAGASAPVLYQGEVTDQKAM 287  
233 -----KPRRLVDSWYD-----AYLGAVLVNIVDGF--- 260

288 GGETSVRENSGIEBKVPSIPKE 311  
261 -----VRKN-----DKVICAQTKK 274

RESULT 9

US-09-134-001C-4705  
 ; Sequence 4705, Application US/09134001C  
 ; Patent No. 6380370  
 ; GENERAL INFORMATION:  
 ; APPLICANT: Lynn Doucette-Stamm et al  
 ; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO STAPHYLOCOCCUS  
 ; TITLE OF INVENTION: EPIDERMIDIS FOR DIAGNOSTICS AND THERAPEUTICS  
 ; FILE REFERENCE: GTC-007  
 ; CURRENT APPLICATION NUMBER: US/09/134,001C  
 ; PRIOR FILING DATE: 1998-08-13  
 ; PRIOR APPLICATION NUMBER: US 60/064,964  
 ; PRIOR FILING DATE: 1997-11-08  
 ; PRIOR APPLICATION NUMBER: US 60/055,779  
 ; PRIOR FILING DATE: 1997-08-14  
 ; NUMBER OF SEQ ID NOS: 5674  
 ; SEQ ID NO 4705  
 ; LENGTH: 540  
 ; TYPE: PRT  
 ; ORGANISM: Staphylococcus epidermidis  
 US-09-134-001C-4705

Query Match 6.3%; Score 101.5; DB 4; Length 540;  
 Best Local Similarity 17.4%; Pred. No. 0.12;  
 Matches 72; Conservative 70; Mismatches 119; Indels 153; Gaps 19;

QY 6 ANPFGCAGVDRAIRIYDQAIKAFAPIVRHEVYVHRTV-----VDGLKQ 51  
 DB 111 ANPVGIRGQIDKAVQVAIEALHISQVENKNEIAQVGAISADEIRGRYISEANDKXGN 170  
 QY 52 KQAVIEELS-----DIPVG-----SYLISAHVSKEVQOE--ABEROLTVFD 93  
 DB 171 DQVITLESNGENTLEVEVEGQPDGQSPFVWVDSKMAELRPIYLVTDKXISSFQ 230  
 QY 94 ATCPITVKVMQVAKAGKQREVILIGHAGPEVEGTQYKTEGGIYIVETPEEVR 153  
 DB 231 DILFLEQV-----VQASRPIIVA--DEVG-----DLAT 259  
 QY 154 NLKAVNP-----NDLAVYTQTLISNTDTKRWADLRQGFPSIK 191  
 DB 260 NIVNRKMGFTPAVAVKAPGFGDRKAMLEDAI-----LTGAQVITDILGLE--LK 309  
 QY 192 EOKKDDICATONRODAVHDLAKISDLILVVGSPSSNSN--RLREIAYQ----- 239  
 DB 310 DASIDML--GIANKVEYTKDHTVVD-----GKDENNIDARVGQIKXQIEETSEDPKE 362  
 QY 240 ----LGRPAVILIDTYQ-----DLK-----QDMLGIEVY--GVTAGASAPVLAQ 278  
 DB 363 KITESIGKLPQGVAVNQGSEETELKERRKLRIEDALNSTRAAVEGIVAGGRTLVNIV 422  
 QY 279 EVIDLTKAMGETTSYR-----ENSGIEEKVVPSTPKELKXMOA 318  
 DB 423 QKVSIRKEDGVEYGVNIVLAKLQAPVROIAENAGLEBSIIV---EELKHAEA 472

## RESULT 10

US-08-467-822-35  
 ; Sequence 35, Application US/08467822  
 ; Patent No. 5843460

; GENERAL INFORMATION:  
 ; APPLICANT: Labigne, Agnes  
 ; APPLICANT: Sauerbaum, Sebastien  
 ; APPLICANT: Ferrero, Richard L.  
 ; APPLICANT: Thibierge, Jean-Michel  
 ; TITLE OF INVENTION: IMMUNOGENIC COMPOSITIONS AGAINST  
 ; TITLE OF INVENTION: HELICOBACTER INFECTION, POLYPEPTIDES FOR USE IN THE  
 ; TITLE OF INVENTION: COMPOSITIONS, AND NUCLEIC ACID SEQUENCES ENCODING SAID  
 ; NUMBER OF SEQUENCES: 44  
 ; CORRESPONDENCE ADDRESS:  
 ; ADDRESSEE: Finnegan, Henderson, Farabow, Garrett &  
 ; STREET: 1300 I Street, N.W.  
 ; CITY: Washington

STATE: D.C.  
 COUNTRY: USA  
 ZIP: 20005-3315  
 COMPUTER READABLE FORM:  
 MEDIUM TYPE: Floppy disk  
 COMPUTER: IBM PC compatible  
 OPERATING SYSTEM: PC-DOS/MS-DOS  
 SOFTWARE: Patent Release #1.0, Version #1.30  
 ; CURRENT APPLICATION DATA:  
 ; APPLICATION NUMBER: US/08/467,822  
 ; FILING DATE: 06-JUN-1995  
 ; CLASSIFICATION: 435  
 ; PRIOR APPLICATION DATA:  
 ; APPLICATION NUMBER: US 08/447,177  
 ; FILING DATE: 19-MAY-1995  
 ; CLASSIFICATION: 435  
 ; PRIOR APPLICATION DATA:  
 ; APPLICATION NUMBER: US 08/432,697  
 ; FILING DATE: 02-MAY-1995  
 ; CLASSIFICATION: 435  
 ; ATTORNEY/AGENT INFORMATION:  
 ; NAME: Meyers, Kenneth J.  
 ; REGISTRATION NUMBER: 25,146  
 ; REFERENCE/DOCKET NUMBER: 03495.0137-02000  
 ; TELECOMMUNICATION INFORMATION:  
 ; TELEPHONE: (202) 408-4000  
 ; TELEFAX: (202) 408-4400  
 ; INFORMATION FOR SEQ ID NO: 35:  
 ; SEQUENCE CHARACTERISTICS:  
 ; LENGTH: 547 amino acids  
 ; TYPE: amino acid  
 ; STRANDEDNESS: single  
 ; TOPOLOGY: linear  
 ; MOLECULE TYPE: protein  
 US-08-467-822-35

Query Match 6.2%; Score 101; DB 2; Length 547;  
 Best Local Similarity 18.2%; Pred. No. 0.14;  
 Matches 72; Conservative 65; Mismatches 143; Indels 116; Gaps 16;

QY 6 ANPFGCAGVDRAIRIYDQAIKAFAPIVRHEVYVHRTV-----VVDGLK-- 50  
 DB 110 ANPVEIRGQVDLAVDAVIAELKQSKVETTPBEIAQVATISANDKEIGNISAMKVG 169  
 QY 51 QKGAFTI-----EELSDVP-----VGSYLIFSAHVSKEVQOE--ABEROLTVF 92  
 DB 170 RKGVIYVKQKTLNDELLEIEGKMFDRGYISPYFINTSKQKCFQDAVYLLSEKXISSI 229  
 QY 93 DATCEPLV-----TKRMOVA-----KAKQREYVILIGHAGPEVEGTMQ 133  
 DB 230 QSVIPALEIANLVNRLKVLGVAVAPGFTLVNRLKVLGVAVAPAPGFGDRKXQQLK 289  
 QY 134 YEKTEGGGIV-----LVETPEVRNLKVNPNNDLAVYTQTLISNTDTKRWVD-----ALR 184  
 DB 290 DMAIATGAVFGEGLTLNEDVQ-----PHDLGKVGIVIVKDDMLLKXGDKXQIE 343  
 QY 185 EOPFSIRKQKDDICATONRODAVHDLAKISD--LILVVGSPSSNSNRLR----- 234  
 DB 344 KRIOEIIIBOL--DVTTSEYEKERNLRIANTLSDGVAVLKVGTSIDVENVKKORVTDALN 401  
 QY 235 ----EIAVOLKPAVL-----IDTYODLKODMLEGIVGVGTAGASAPVLAQEVVD 282  
 DB 402 ATRAAVEGIVLGGGCLLRICFALDSLTPTANEDQKIGIETIRTKLIPA----- 451  
 QY 283 QLRWGGETTSVRENSGIEEKVVPSTPKELKXMOA 318  
 DB 452 ----MTIAKNAGVDSLSI-----VEKIMQS 472

## RESULT 11

US-08-432-697-35  
 ; Sequence 35, Application US/08432697  
 ; Patent No. 6248330

GENERAL INFORMATION:  
APPLICANT: Labigne, Agnes  
APPLICANT: Sauerbaum, Sebastien  
APPLICANT: Petro, Richard L.  
APPLICANT: Thibierge, Jean-Michel  
TITLE OF INVENTION: IMMUNOGENIC COMPOSITIONS AGAINST  
TITLE OF INVENTION: HELICOBACTER INFECTION, POLYPEPTIDES FOR USE IN THE  
TITLE OF INVENTION: COMPOSITIONS, AND NUCLEIC ACID SEQUENCES ENCODING SAID  
NUMBER OF INVENTION: POLYPEPTIDES  
NUMBER OF SEQUENCES: 44  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: Fimegan, Henderson, Farabow, Garrett &  
ADDRESS: Dunner  
STREET: 1300 I Street, N.W.  
CITY: Washington  
STATE: D.C.  
COUNTRY: USA  
ZIP: 20005-3315  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: Patentin Release #1.0, Version #1.30  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/432,697  
FILING DATE: 02-MAY-1995  
CLASSIFICATION: 424  
ATTORNEY/AGENT INFORMATION:  
NAME: Meyers, Kenneth J.  
REGISTRATION NUMBER: 25,146  
REFERENCE/DOCKET NUMBER: 03495, 0137-00000  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: (202) 408-4400  
TELEFAX: (202) 408-4400  
INFORMATION FOR SEQ ID NO: 35:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 547 amino acids  
TYPE: amino acid  
STRANDEDNESS: single  
TOPOLOGY: linear  
MOLECULE TYPE: protein  
US-08-432-697-35

Query Match 6.2%; Score 101; DB 3; Length 547;  
Best Local Similarity 18.2%; Pred. No. 0.14;  
Matches 72; Conservative 65; Mismatches 143; Indels 116; Gaps 16;

6 ANRPGFCAGVRAIRIVDAIAFGAPIYVRHEVHNRT-----VVDGLK--- 50  
110 ANPVEIRRGVDLAVVAIAELKQSKPTTPEIQAQVATISANGKEIGNISDMKKVG 169  
51 QKGAVF-----BELSDVP-----VGSYLIFSANGSVKVOOE--AEERQLTVF 92  
170 RKGVTITVDKGTINDELITIBGKFDGVIISPYFINTSKQKCEPDQAVVLSKKISST 229  
93 DATCPV-----TKYHQVA-----KGAQGRREVLILIGHAPREVBGMQ 133  
230 QSIAPALAILANLVNRLKGLQVAVKAPGLVLANRLKGLQVAVKAPGSGDNRKQOLK 289  
134 YEKTEGGGIVY---LVETPEDVRNLKVNPNPDIAVYVQTLSMTDRTYVWD-----ALR 184  
230 DMALITGGAVFGEELGTINLEEDVQ-----PHDLQKVEEIVITKQDAMLKGGKQKQITE 343  
185 EGFPSIKQKKODICVATQNRQDAVHDLAKISD--LIIVGSPNSNSNRRL----- 234  
344 KRIQIITQL--DVTTSEYKSKLNERLAKSDGVAVLKVGTSVVEVNEKKDQVTDALN 401  
235 -----ELAVOLGKPAVL-----IDTYQDLKQDMLBEGIEVVGVTAGASAPVLEVOEYID 282  
402 ATRAAVEGIVAGGCALRCIPALDSTLPANEDQKIGIITIKRLKIPA----- 451  
283 QLKMGGETTSVRENSGIEEKVFSIKELKQKQA 318

DB 452-----MTIAKAGVDSGLI-----VEKIMQS 472

RESULT 12  
US-08-466-248-35  
Sequence 35, Application US/08466248  
Patent No. 6258359  
GENERAL INFORMATION:  
APPLICANT: Labigne, Agnes  
APPLICANT: Sauerbaum, Sebastien  
APPLICANT: Petro, Richard L.  
APPLICANT: Thibierge, Jean-Michel  
TITLE OF INVENTION: IMMUNOGENIC COMPOSITIONS AGAINST  
TITLE OF INVENTION: HELICOBACTER INFECTION, POLYPEPTIDES FOR USE IN THE  
TITLE OF INVENTION: COMPOSITIONS, AND NUCLEIC ACID SEQUENCES ENCODING SAID  
NUMBER OF SEQUENCES: 44  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: Fimegan, Henderson, Farabow, Garrett &  
ADDRESS: Dunner  
STREET: 1300 I Street, N.W.  
CITY: Washington  
STATE: D.C.  
COUNTRY: USA  
ZIP: 20005-3315  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: Patentin Release #1.0, Version #1.30  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/466,248  
FILING DATE: 06-JUN-1995  
CLASSIFICATION: 435  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US 08/447,177  
FILING DATE: 19-MAY-1995  
CLASSIFICATION: 435  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US 08/432,697  
FILING DATE: 02-MAY-1995  
CLASSIFICATION: 435  
ATTORNEY/AGENT INFORMATION:  
NAME: Meyers, Kenneth J.  
REGISTRATION NUMBER: 25,146  
REFERENCE/DOCKET NUMBER: 03495, 0137-02000  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: (202) 408-4400  
TELEFAX: (202) 408-4400  
INFORMATION FOR SEQ ID NO: 35:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 547 amino acids  
TYPE: amino acid  
STRANDEDNESS: single  
TOPOLOGY: linear  
MOLECULE TYPE: protein  
US-08-466-248-35

Query Match 6.2%; Score 101; DB 3; Length 547;  
Best Local Similarity 18.2%; Pred. No. 0.14;  
Matches 72; Conservative 65; Mismatches 143; Indels 116; Gaps 16;

6 ANRPGFCAGVRAIRIVDAIAFGAPIYVRHEVHNRT-----VVDGLK--- 50  
110 ANPVEIRRGVDLAVVAIAELKQSKPTTPEIQAQVATISANGKEIGNISDMKKVG 169  
51 QKGAVF-----BELSDVP-----VGSYLIFSANGSVKVOOE--AEERQLTVF 92  
170 RKGVTITVDKGTINDELITIBGKFDGVIISPYFINTSKQKCEPDQAVVLSKKISST 229  
93 DATCPV-----TKYHQVA-----KGAQGRREVLILIGHAPREVBGMQ 133  
230 QSIAPALAILANLVNRLKGLQVAVKAPGLVLANRLKGLQVAVKAPGSGDNRKQOLK 289

QY 134 YEKTEGGGIX-----LVETPEDVRNKKNNPNDAVYTTTSTMTDKYMD-----ALR 184  
 Db 290 DWAATGAVFGEHGTTLNLEDO-----PHDLGVGEVITVKDAMLLKKGKDAQIE 343  
 QY 185 EOPPSIKEOKKDDICATONRODAVHDLAKISD--LIVVGSPPNSNSNRLLR----- 234  
 Db 344 KRIQIEIIEQL--DVTSEYKKEKLNRLAKLSGVAVLKVGGSIDVYVWEKKDRVTDALN 401  
 QY 235 -----EIVVQKSPVY-----IDTYODLKQDLLEGIEVGVTAAGASAPVLYOEVLD 282  
 Db 402 ATRAAVEEIVGGGACALRCIPALDSLPANEDOKIGIEIIRKTLKJPA----- 451  
 QY 283 OLKAWGSETTSVENGEIEKVVFSIPKELKKQMOA 318  
 Db 452 -----MTAKNAGVUGSLI-----VEKIMOS 472

RESULT 13  
 US-08-144-121-4  
 ; Sequence 4, Application US/08144121  
 ; Patent No. 5610031  
 ; GENERAL INFORMATION:  
 ; APPLICANT: Burgeson, Robert E.  
 ; TITLE OF INVENTION: BIX CHAIN OF LAMININ AND METHODS OF USE  
 ; NUMBER OF SEQUENCES: 14  
 ; CORRESPONDENCE ADDRESS:  
 ; ADDRESSEE: LAHIVE & COCKFIELD  
 ; STREET: 60 State Street, suite 510  
 ; CITY: BOSTON  
 ; STATE: Massachusetts  
 ; COUNTRY: United States  
 ; ZIP: 02109  
 ; COMPUTER READABLE FORM:  
 ; MEDIUM TYPE: Floppy disk  
 ; COMPUTER: IBM PC compatible  
 ; OPERATING SYSTEM: PC-DOS/MS-DOS  
 ; SOFTWARE: Patentin Release #1.0, Version #1.25  
 ; CURRENT APPLICATION DATA:  
 ; APPLICATION NUMBER: US/08/144,121  
 ; FILING DATE: 27-OCT-1993  
 ; CLASSIFICATION: 424  
 ; ATTORNEY/AGENT INFORMATION:  
 ; NAME: Myers, Paul L.  
 ; REGISTRATION NUMBER: 35,965  
 ; REFERENCE/DOCKET NUMBER: (MGP-0780.0) MGP-021  
 ; TELECOMMUNICATION INFORMATION:  
 ; TELEPHONE: (617) 227-7400  
 ; TELEFAX: (617) 227-5941  
 ; INFORMATION FOR SEQ ID NO: 4:  
 ; SEQUENCE CHARACTERISTICS:  
 ; LENGTH: 1196 amino acids  
 ; TYPE: amino acid  
 ; TOPOLOGY: linear  
 ; MOLECULE TYPE: peptide  
 ; FRAGMENT TYPE: internal  
 ; FEATURE:  
 ; NAME/KEY: Domain  
 ; LOCATION: 1..250  
 ; FEATURE:  
 ; NAME/KEY: Domain  
 ; LOCATION: 251..437  
 ; FEATURE:  
 ; NAME/KEY: Domain  
 ; LOCATION: 438..807  
 ; FEATURE:  
 ; NAME/KEY: Domain  
 ; LOCATION: 808..840  
 ; FEATURE:  
 ; NAME/KEY: Domain  
 ; LOCATION: 841..1196  
 ; US-08-144-121-4

Query Match 6.1%; Score 99.5; DB 1; Length 1196;  
 Best Local Similarity 19.8%; Freq. No. 0.71;  
 Matches 52; Conservative 67; Mismatches 91; Indels 53; Gaps 13;  
 QY 69 LIPSAHGV-----SKRQOE-----AEERQLTFDTCPLVTKVHMVQVAKAQGEVILIG 120  
 Db 843 LVYAHNNMQAMLDQDVLSALAEVEQLS-----KRVSAKLR-ADBAQSAEDILL- 894  
 QY 121 HAHPEVEGTGQYIEKTEGGGIVLVETPEDVRNKKNNPNDAVYTTTSTMTDKV 178  
 Db 895 -----KTNATEKRNKK-----SNEELRNILIXQIRN-----FLTQSDADLDSTEA 933  
 QY 179 MV-DALREQPSIEOKKDDICATONRODAVHDLAKISD--LIVVGSPPNSNSNRLLRIA 237  
 Db 934 VANEVLKKNMSTFQQLQN-----LTEDIRERESLSQV-EVILQHSADIRAEKTLERA 988  
 QY 238 VQKGPVYLIDTYODLKQDLLEGIEVGVTAAGASAPVLYOEVLDOLKAWGSETTSVEN 297  
 Db 989 KRASKSATDVKVTADVWEALAEAKQVAA-----EKAIQADEDIGTQNLNLTSSBS 1043  
 QY 298 SGIEKVVFSIPK--BLKKMQ 317  
 Db 1044 TAASETLFFNASORISBLERVVE 1066

RESULT 14  
 US-08-735-893-4  
 ; Sequence 4, Application US/08735893  
 ; Patent No. 5914317  
 ; GENERAL INFORMATION:  
 ; APPLICANT: Burgeson, Robert E.  
 ; TITLE OF INVENTION: BIX CHAIN OF LAMININ AND METHODS OF USE  
 ; NUMBER OF SEQUENCES: 14  
 ; CORRESPONDENCE ADDRESS:  
 ; ADDRESSEE: LAHIVE & COCKFIELD  
 ; STREET: 60 State Street, suite 510  
 ; CITY: BOSTON  
 ; STATE: Massachusetts  
 ; COUNTRY: United States  
 ; ZIP: 02109  
 ; COMPUTER READABLE FORM:  
 ; MEDIUM TYPE: Floppy disk  
 ; COMPUTER: IBM PC compatible  
 ; OPERATING SYSTEM: PC-DOS/MS-DOS  
 ; SOFTWARE: Patentin Release #1.0, Version #1.25  
 ; CURRENT APPLICATION DATA:  
 ; APPLICATION NUMBER: US/08/735,893  
 ; FILING DATE: 18-OCT-1996  
 ; CLASSIFICATION: 435  
 ; PRIOR APPLICATION NUMBER:  
 ; APPLICATION NUMBER: US 08/144,121  
 ; FILING DATE: 27-OCT-1993  
 ; CLASSIFICATION: 435  
 ; ATTORNEY/AGENT INFORMATION:  
 ; NAME: Myers, Paul L.  
 ; REGISTRATION NUMBER: 35,965  
 ; REFERENCE/DOCKET NUMBER: (MGP-0780.1) MGP-021DV  
 ; TELECOMMUNICATION INFORMATION:  
 ; TELEPHONE: (617) 227-7400  
 ; TELEFAX: (617) 227-5941  
 ; INFORMATION FOR SEQ ID NO: 4:  
 ; SEQUENCE CHARACTERISTICS:  
 ; LENGTH: 1196 amino acids  
 ; TYPE: amino acid  
 ; TOPOLOGY: linear  
 ; MOLECULE TYPE: peptide  
 ; FRAGMENT TYPE: internal  
 ; FEATURE:  
 ; NAME/KEY: Domain  
 ; LOCATION: 1..250  
 ; FEATURE:

NAME/KEY: Domain  
 LOCATION: 251..437  
 FEATURE:  
 NAME/KEY: Domain  
 LOCATION: 438..807  
 FEATURE:  
 NAME/KEY: Domain  
 LOCATION: 808..840  
 FEATURE:  
 NAME/KEY: Domain  
 LOCATION: 841..1196  
 US-08-797-3583-4

Query Match 6.1%; Score 99.5; DB 2; Length 1196;  
 Best Local Similarity 19.8%; Pred. No. 0.71;

Matches 52; Conservative 67; Mismatches 91; Indels 53; Gaps 13;

QY 69 LIFSAGV---SKEVQOE---AERQUTVPDAPLVTKVHQAQAGREVILIG 120  
 DB 843 LVTVAHNAWQKAMDIDQVLSALAVEQLS-----KQVSEAKLR-ADEAKQSAEDILL- 894  
 QY 121 HAGHEVEGTMGQVCKTGGGIVLVETPEDEVRL--KVNPNDAVYVTTLSMTDIXV 178  
 DB 895 -----KTNATKEKMK-----SNEELRNLIKQIRN-----FLTQSDADLSIEA 933  
 QY 179 MV-DALREQPSIKEQKDDICATONRODAVHDAKISDILVYGSNSNSNRLEIA 237  
 DB 934 VANETLKNMESTPQOLQV---LTEDIRBRVESLSQV-EVILQHSADIAHAEVLLIEA 988  
 QY 238 VOLGPAVILDTYQDLKQDMLEIGIVVGTAGASAPENVQEVIDQLKAMGETTSVREN 297  
 DB 969 KRASSATADVKTADVMKEALEBAKQAVNA-----EKAQKQADEIDQGTQMLTISSE 1043  
 QY 238 SGIEKVFVESIPK---ELKQHQ 317  
 DB 1044 TRASETELENSQRISELRNVE 1066

RESULT 15  
 US-08-797-358B-3

; Sequence 3, Application US/08797358B  
 ; Patent No. 6268478

; GENERAL INFORMATION:

; APPLICANT: Adams, John

; TITLE OF INVENTION: INTRACELLULAR VITAMIN D BINDING PROTEIN

; NUMBER OF SEQUENCES: 9

; CORRESPONDENCE ADDRESS:

; ADDRESSEE: Campbell & Flores LLP

; STREET: 4370 La Jolla Village Drive, Suite 700

; CITY: San Diego

; STATE: California

; COUNTRY: United States

; ZIP: 92122

; COMPUTER READABLE FORM:

; MEDIUM TYPE: Floppy disk

; COMPUTER: IBM PC compatible

; OPERATING SYSTEM: PC-DOS/MS-DOS

; SOFTWARE: Patent Release #1.0, Version #1.25

; CURRENT APPLICATION NUMBER: US/08/797,358B

; FILING DATE: 11-Feb-1997

; CLASSIFICATION: <Unknown>

; PRIOR APPLICATION DATA:

; APPLICATION NUMBER: US 60/011,491

; FILING DATE: 12-FEB-1996

; ATTORNEY/AGENT INFORMATION:

; NAME: Campbell, Cathryn A.

; REGISTRATION NUMBER: 31,815

; REFERENCE/DOCKET NUMBER: P-CE 3165

; TELECOMMUNICATION INFORMATION:

; TELEPHONE: (619) 535-9001

; TELEFAX: (619) 535-8949

; INFORMATION FOR SEQ ID NO: 3:

SEQUENCE CHARACTERISTICS:  
 LENGTH: 643 amino acids  
 TYPE: amino acid  
 STRANDEDNESS: unknown  
 TOPOLOGY: unknown  
 MOLECULE TYPE: protein  
 SEQUENCE DESCRIPTION: SEQ ID NO: 3:  
 US-08-797-358B-3

Query Match 5.9%; Score 95.5; DB 3; Length 643;

Best Local Similarity 21.9%; Pred. No. 0.67;

Matches 77; Conservative 45; Mismatches 102; Indels 127; Gaps 18;

QY 24 QATEAF-GAITYVRHEV-----HNRIVDGLKQKAVF---IEELSDVVGSLIFSAG 74  
 DB 131 EYBAVILGQP--VHAATVPAYFNDSQROATDAGATVGLAVLRINERTAAAI--AH 185  
 QY 75 GVSKEVQQAERQUTVFD--ATCPVTKVHQAQAGREVILIGHAGHEVEGTM 131  
 DB 186 GLDR--RGAGERNVILFDLGGGTFDV-----SVLSIDAGVEVAKATA 225  
 QY 132 GQYKCTGSGGIVLVETPEDEVRLKYN-----NPNDAV-----VTQ 168  
 DB 226 GD-----THLOG-----KDPDRNLVNHFEVERKRRKRLSNKRALRLRTACERAKR 274  
 QY 169 TTLSMTDTRVVDALREQ---PSIKEQKDDICV-----ATONRODAVHDLAKIS 216  
 DB 275 TLSSSTQATLEIDSLFEGVVPYTSITPAKREELCSDLFRSTLEPVEKGLADKXKXIH 334  
 QY 217 DLILVGSFNSSNSNRLEIAVQLGKPAVILDTYQDLKQDMLEIGIEV-----VGVTAG 269  
 DB 335 DVVLVGSSTR-----IPRVQKLIQDFPFGKELINKSINPDEAVAYG 374  
 QY 270 ASAPEVAV-----QEVTDOL-----YAMGETTSVRENSGIEEK 303  
 DB 375 AAVQAIVLMGDKCKKVDLILLDVAPLSLGLERAGVMTLLIQRNATIPTR 425

Search completed: January 29, 2004, 15:57:09  
 Job time : 11.9804 secs

GenCore version 5.1.6  
Copyright (c) 1993 - 2004 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: January 29, 2004, 15:54:19 : Search time 21.7073 Seconds

(without alignments)  
3044.503 Million cell updates/sec

Title: US-09-941-947a-18

Perfect score: 1623  
Sequence: 1 MQVLANPGRGCAVDRAIE.....GIEKVVPSIKELKKHQA 318

Scoring table: BLOSUM62  
Gapop 10.0 , Gapext 0.5

Searched: 789580 seqs, 207824079 residues

Total number of hits satisfying chosen parameters: 789580

Minimum DB seq length: 0  
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Database :

Listing first 45 summaries

Published Applications AA:\*

1: /cgn2\_6/ptodata/1/pubpaa/US07\_PUBCOMB.pep:\*

2: /cgn2\_6/ptodata/1/pubpaa/PC1\_NEM\_PUB.pep:\*

3: /cgn2\_6/ptodata/1/pubpaa/US06\_NEM\_PUB.pep:\*

4: /cgn2\_6/ptodata/1/pubpaa/US06\_PUBCOMB.pep:\*

5: /cgn2\_6/ptodata/1/pubpaa/US07\_NEM\_PUB.pep:\*

6: /cgn2\_6/ptodata/1/pubpaa/PC07\_PUBCOMB.pep:\*

7: /cgn2\_6/ptodata/1/pubpaa/US08\_NEM\_PUB.pep:\*

8: /cgn2\_6/ptodata/1/pubpaa/US08\_PUBCOMB.pep:\*

9: /cgn2\_6/ptodata/1/pubpaa/US09\_PUBCOMB.pep:\*

10: /cgn2\_6/ptodata/1/pubpaa/US09C\_PUBCOMB.pep:\*

11: /cgn2\_6/ptodata/1/pubpaa/US09C\_PUBCOMB.pep:\*

12: /cgn2\_6/ptodata/1/pubpaa/US09\_NEM\_PUB.pep:\*

13: /cgn2\_6/ptodata/1/pubpaa/US10A\_PUBCOMB.pep:\*

14: /cgn2\_6/ptodata/1/pubpaa/US10B\_PUBCOMB.pep:\*

15: /cgn2\_6/ptodata/1/pubpaa/US10C\_PUBCOMB.pep:\*

16: /cgn2\_6/ptodata/1/pubpaa/US10C\_NEM\_PUB.pep:\*

17: /cgn2\_6/ptodata/1/pubpaa/US60\_NEM\_PUB.pep:\*

18: /cgn2\_6/ptodata/1/pubpaa/US60\_PUBCOMB.pep:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

#### SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	1623	100.0	318	10	US-09-934-903-24
2	1623	100.0	318	11	US-09-941-947a-18
3	1028.5	63.4	349	11	US-09-975-719-429
4	787	48.5	338	15	US-10-156-761-10747
5	755	46.5	325	10	US-09-738-626-4643
6	743.5	45.2	329	10	US-09-712-363-271
7	733	45.8	335	10	US-09-712-363-188
8	663	40.9	310	12	US-10-289-762-1090
9	663	40.9	310	15	US-10-022-832-16
10	598.5	36.9	539	15	US-10-275-360-5
11	111	6.8	917	15	US-10-156-761-10047
12	109	6.7	578	10	US-09-712-363-282
13	109	6.7	925	12	US-10-369-493-5415
14	104.5	6.4	501	12	US-10-369-493-1052
15	104.5	6.4	1130	12	US-10-369-493-19263

16	104	6.4	869	12	US-10-369-493-19983	Sequence 19983, A
17	103	6.3	433	9	US-09-815-242-10607	Sequence 10607, A
18	103	6.3	553	12	US-10-032-585-7338	Sequence 7338, Ap
19	102	6.3	1382	10	US-09-729-653-2	Sequence 2, Appl1
20	101	6.2	391	12	US-10-369-493-20969	Sequence 20969, A
21	101	6.2	520	12	US-10-369-493-2893	Sequence 2893, Ap
22	100.5	6.2	538	12	US-10-369-493-2950	Sequence 2950, Ap
23	99.5	6.1	1765	15	US-10-037-182-8	Sequence 8, Appl1
24	99.5	6.1	1786	9	US-08-873-676-113	Sequence 113, App
25	99.5	6.1	1786	9	US-09-938-275-6	Sequence 6, Appl1
26	99.5	6.1	1786	15	US-10-037-182-6	Sequence 6, Appl1
27	97.5	6.0	607	12	US-10-369-493-19473	Sequence 19473, A
28	97.5	6.0	921	10	US-09-738-626-5659	Sequence 5659, Ap
29	97	6.0	299	12	US-10-369-493-10700	Sequence 10700, A
30	96.5	5.9	321	9	US-09-815-242-13454	Sequence 14570, A
31	96.5	5.9	1881	12	US-10-032-585-7646	Sequence 7646, Ap
32	96	5.9	952	12	US-10-108-260A-3284	Sequence 3284, Ap
33	95.5	5.9	330	12	US-10-369-493-23413	Sequence 23413, A
34	95.5	5.9	977	12	US-10-342-136-2	Sequence 2, Appl1
35	94.5	5.8	759	12	US-10-369-493-18629	Sequence 18629, A
36	94	5.8	496	15	US-10-215-224-7	Sequence 7, Appl1
37	94	5.8	496	15	US-10-263-677-9	Sequence 9, Appl1
38	94	5.8	573	10	US-09-847-637B-8	Sequence 8, Appl1
39	93.5	5.8	547	12	US-10-369-493-7917	Sequence 7917, Ap
40	93	5.7	563	12	US-10-341-434-135	Sequence 135, App
41	93	5.7	573	10	US-09-828-574-1	Sequence 1, Appl1
42	93	5.7	573	12	US-10-307-326-1	Sequence 1, Appl1
43	93	5.7	573	12	US-10-116-275-141	Sequence 141, App
44	93	5.7	573	12	US-10-367-093-12	Sequence 12, Appl1
45	93	5.7	573	15	US-10-046-649-1	Sequence 1, Appl1

#### ALIGNMENTS

RESULT 1

US-09-934-903-24

Sequence 24, Application US/09934903

Patent No. US20020102690A1

GENERAL INFORMATION:

APPLICANT: Koffas, Mattheos

APPLICANT: Odum, J. Martin

APPLICANT: Schenckle, Andreas J.

APPLICANT: No. US20020102690A1ton, Kelley C.

APPLICANT: Tomb, Jean-Francois

APPLICANT: Rouviere, Pierre

APPLICANT: Picataggio, Stephen

APPLICANT: Cheng, Qiong

TITLE OR INVENTION: Genes Involved in Isoprenoid Compounds Production

FILE REFERENCE: C11646 US NA

CURRENT APPLICATION NUMBER: US/09/934,903

CURRENT FILING DATE: 2001-08-22

PRIOR APPLICATION NUMBER: 60/229,907

PRIOR FILING DATE: September 1, 2001

NUMBER OF SEQ ID NOS: 24

SOFTWARE: Microsoft Office 97

SEQ ID NO 24

LENGTH: 318

TYPE: PRT

ORGANISM: Methylobionas 16a

US-09-934-903-24

Query Match 100.0%; Score 1623; DB 10; Length 318;

Best local Similarity 100.0%; Pred. No. 2.5e-150; Indels 0; Gaps 0;

Matches 318; Conservative 0; Mismatches 0;

DB

QY 1 MQVLANPGRGCAVDRAIEIVDAIFAFGAPIVRRHVHNRITVVDGLKQKGAFTBEI 60

DB 1 MQVLANPGRGCAVDRAIEIVDAIFAFGAPIVRRHVHNRITVVDGLKQKGAFTBEI 60

QY 61 SDPVVSGYLLIFSANGVKEVQQAERQLTYFDATGCHLYTKVHQAQKAGREVLILIG 120

DB 61 SDPVVSGYLLIFSANGVKEVQQAERQLTYFDATGCHLYTKVHQAQKAGREVLILIG 120

```

QY 121 HGHPEVEGTMGQYEKCTGGGIIYVETPEDEVRLKVNPNNDLAYVTTLSMTDTKVM 180
DB 121 HGHPEVEGTMGQYEKCTGGGIIYVETPEDEVRLKVNPNNDLAYVTTLSMTDTKVM 180
QY 181 DALREOPPSIKQKODICYATQNRDADVHDLAKISDLILVVGSPNSNSRLREIAVOL 240
DB 181 DALREOPPSIKQKODICYATQNRDADVHDLAKISDLILVVGSPNSNSRLREIAVOL 240
QY 241 GKPAYLIDTYODLKODMLBGIEVGVTAASAPVILVQEVITDOLKAWGGETTSVRENSGI 300
DB 241 GKPAYLIDTYODLKODMLBGIEVGVTAASAPVILVQEVITDOLKAWGGETTSVRENSGI 300
QY 301 EEKVVPISIPKELKXKMOA 318
DB 301 EEKVVPISIPKELKXKMOA 318

```

## RESULT 2

```

US-09-941-947A-18
; Sequence 18, Application US/09941947A
; Publication No. US20030003528A1
; GENERAL INFORMATION:
; APPLICANT: Brzostowicz, Patricia C.
; APPLICANT: Cheng, Qiong
; APPLICANT: Dicosimo, Deana J.
; APPLICANT: Koffas, Matthews
; APPLICANT: Miller, Edward S. Jr.
; APPLICANT: Ogden, J. Martin
; APPLICANT: Picatoggio, Steve
; APPLICANT: Rouviere, Pierre B.
; TITLE OF INVENTION: CAROTENOID PRODUCTION FROM A SINGLE CARBON SOURCE
; FILE REFERENCE: CL1903 US NA
; CURRENT APPLICATION NUMBER: US/09/941,947A
; CURRENT FILING DATE: 2001-09-01
; PRIOR APPLICATION NUMBER: 60/229,907
; PRIOR FILING DATE: 2000-09-01
; PRIOR APPLICATION NUMBER: 60/229,858
; PRIOR FILING DATE: 2000-09-01
; NUMBER OF SEQ ID NOS: 60
; SOFTWARE: Microsoft Office 97
; SEQ ID NO 18
; LENGTH: 318
; TYPE: PRT
; ORGANISM: Methylomonas 16a
; US-09-941-947A-18

```

```

Query Match 100.0%; Score 1623; DB 11; Length 318;
Best Local Similarity 100.0%; Pred. No. 2.5e-150;
Matches 318; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```

```

QY 1 MOIVLANPRGFCAGVDRAIEIVDAIEAFGADIVYRHEVHNRTVVDGLKQGAFFIEEL 60
DB 1 MOIVLANPRGFCAGVDRAIEIVDAIEAFGADIVYRHEVHNRTVVDGLKQGAFFIEEL 60
QY 61 SDVPVGSYILFSAGVSKXVOQAEEROLTVFPATCPLVTVMQVAKAKQGEVILLIG 120
DB 61 SDVPVGSYILFSAGVSKXVOQAEEROLTVFPATCPLVTVMQVAKAKQGEVILLIG 120
QY 121 HGHPEVEGTMGQYEKCTGGGIIYVETPEDEVRLKVNPNNDLAYVTTLSMTDTKVM 180
DB 121 HGHPEVEGTMGQYEKCTGGGIIYVETPEDEVRLKVNPNNDLAYVTTLSMTDTKVM 180
QY 181 DALREOPPSIKQKODICYATQNRDADVHDLAKISDLILVVGSPNSNSRLREIAVOL 240
DB 181 DALREOPPSIKQKODICYATQNRDADVHDLAKISDLILVVGSPNSNSRLREIAVOL 240
QY 241 GKPAYLIDTYODLKODMLBGIEVGVTAASAPVILVQEVITDOLKAWGGETTSVRENSGI 300
DB 241 GKPAYLIDTYODLKODMLBGIEVGVTAASAPVILVQEVITDOLKAWGGETTSVRENSGI 300
QY 301 EEKVVPISIPKELKXKMOA 318
DB 301 EEKVVPISIPKELKXKMOA 318

```

```

DB 301 EEKVVPISIPKELKXKMOA 318

```

## RESULT 3

```

US-09-975-719-429
; Sequence 429, Application US/09975719
; Publication No. US20030022349A1
; GENERAL INFORMATION:
; APPLICANT: Ausubel, Frederick M.
; APPLICANT: Rahme, Laurence G.
; TITLE OF INVENTION: VIRULANCE-ASSOCIATED NUCLEIC ACID
; FILE REFERENCE: 00786/361003
; CURRENT APPLICATION NUMBER: US/09/975,719
; CURRENT FILING DATE: 2001-10-10
; PRIOR APPLICATION NUMBER: US 09/199,637
; PRIOR FILING DATE: 1998-11-25
; PRIOR APPLICATION NUMBER: US 60/066,517
; PRIOR FILING DATE: 1997-11-25
; NUMBER OF SEQ ID NOS: 437
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 429
; LENGTH: 349
; TYPE: PRT
; ORGANISM: Pseudomonas aeruginosa
; US-09-975-719-429

```

```

Query Match 63.4%; Score 1026.5; DB 11; Length 349;
Best Local Similarity 63.3%; Pred. No. 4.7e-92;
Matches 198; Conservative 49; Mismatches 63; Indels 3; Gaps 2;

```

```

QY 1 MOIVLANPRGFCAGVDRAIEIVDAIEAFGADIVYRHEVHNRTVVDGLKQGAFFIEEL 60
DB 36 MOIVLANPRGFCAGVDRAIEIVDAIEAFGADIVYRHEVHNRTVVDGLKQGAFFIEEL 95
QY 61 SDVPVGSYILFSAGVSKXVOQAEEROLTVFPATCPLVTVMQVAKAKQGEVILLIG 120
DB 96 SDVPVGSYILFSAGVSKXVOQAEEROLTVFPATCPLVTVMQVAKAKQGEVILLIG 155
QY 121 HGHPEVEGTMGQYEKCTGGGIIYVETPEDEVRLKVNPNNDLAYVTTLSMTDTKVM 180
DB 121 HGHPEVEGTMGQYEKCTGGGIIYVETPEDEVRLKVNPNNDLAYVTTLSMTDTKVM 180
QY 156 HGHPEVEGTMGQYEKCTGGGIIYVETPEDEVRLKVNPNNDLAYVTTLSMTDTKVM 214
DB 156 HGHPEVEGTMGQYEKCTGGGIIYVETPEDEVRLKVNPNNDLAYVTTLSMTDTKVM 214
QY 181 DALREOPPSIKQKODICYATQNRDADVHDLAKISDLILVVGSPNSNSRLREIAVOL 240
DB 215 DALREOPPSIKQKODICYATQNRDADVHDLAKISDLILVVGSPNSNSRLREIAVOL 274
QY 241 GKPAYLIDTYODLKODMLBGIEVGVTAASAPVILVQEVITDOLKAWGGETTSVRENSGI 300
DB 275 GKPAYLIDTYODLKODMLBGIEVGVTAASAPVILVQEVITDOLKAWGGETTSVRENSGI 332
QY 301 EEKVVPISIPKELK 313
DB 333 EENITFSPKELR 345

```

## RESULT 4

```

US-10-156-761-10747
; Sequence 10747, Application US/10156761
; Publication No. US20030119018A1
; GENERAL INFORMATION:
; APPLICANT: OMDRA, SATOSHI
; APPLICANT: ISHIZAKA, JUN
; APPLICANT: HORIKAWA, HIROSHI
; APPLICANT: SHIBA, TADAYOSHI
; APPLICANT: SAKAKI, YOSHIYUKI
; APPLICANT: HATTORI, MASAHIRA
; TITLE OF INVENTION: NOVEL POLYNUCLEOTIDES
; FILE REFERENCE: 249-262
; CURRENT APPLICATION NUMBER: US/10/156,761
; CURRENT FILING DATE: 2002-05-29
; PRIOR APPLICATION NUMBER: JP 2001-204089

```

? PRIOR FILING DATE: 2001-05-30  
 ? PRIOR APPLICATION NUMBER: JP 2001-272697  
 ? PRIOR FILING DATE: 2001-08-02  
 ? NUMBER OF SEQ. ID NOS: 15109  
 ? SEQ. ID NO. 10747  
 ? LENGTH: 338  
 ? TYPE: PR1  
 ? ORGANISM: Streptomyces avermitilis  
 ? US-10-156-761-10747

Query Match	48.5%;	Score 787;	DB 15;	Length 338;
Best Local Similarity	48.1%;	Pred. No. 2e-68;		
Matches 153;	Conservative 71;	Mismatches 86;	Indels 8;	Gaps 3

```

QY QIVLAPRGGCAGVDAIILIVQALBAPCAPTYVHEVHNRVTVDGKQAGAFIELS 61
Db 11 RVLAPRGGCAGVDAIVAEKALBOYCAPTYVHEIVHNRVYQTLERGAIVERTA 70
QY 62 DVPVSGYLIFSAHGVSEKVOQEAERKQLVPAATLPTKVMQAKHAKGKRGREVLIGH 122
Db 71 EYVEBGATVFSBAGVAPVTHEBAPNGKATTDATLPTKVKHEKVRFRANDPFIILIGH 130
QY 122 AGHPEVEGTMOYKECTBGGIYLVETBEDVENEKVNPNDLAVYVQTLSMTETKVMVD 181
Db 131 EGHHEEYIGTSGEAPR-----HITLVDPGPDVAKVEVRDPSKVMWLSQTTLSVDTEMETVD 189
QY 182 ALRBPSPSIRKEKKDICYATQNRDVAHDLAKISDILLVGSPKSSNSNRLEJANVLG 244
Db 186 ALRKEKFPQLSPSPSDICYATQNRQILANKQMEADLIVGSRSSNSVATLVAAVLG 245
QY 242 -KRAYLIDPYODLKQPMLEGEIVGVGTAGASAPYLVOEVIDOLKAMGGETTSFRENIGI 300
Db 246 ARDAYLVDPADRIDEMALMGVSTGVGTASAPYELLVQVLEMTLSGRFDELVEYK--AA 303
QY 301 BEKVVSFIRKELKKKKA 318
Db 304 BBSITFSLEKELARDLRA 321

```

```

US RESULT 5
US-09-738-626-4643
Sequence 4643, Application US/09738626
Publication No. US20020197605A1
GENERAL INFORMATION:
APPLICANT: NAKAGAWA, SATOSHI
APPLICANT: MIZOGUCHI, HIROSHI
APPLICANT: ANDO, SEIKO
APPLICANT: HAYASHI, MIKIRO
APPLICANT: YOKOI, HARUHIRO
APPLICANT: TATEISHI, NAOKO
APPLICANT: SENDO, AKIHIRO
APPLICANT: IKEDA, MASATO
APPLICANT: OZAKI, AKIO
TITLE OF INVENTION: NOVEL POLYNUCLEOTIDES
FILE REFERENCE: 249-125
CURRENT APPLICATION NUMBER: US/09/738,626
CURRENT FILING DATE: 2000-12-18
PRIOR APPLICATION NUMBER: JP 99/377484
PRIOR FILING DATE: 1999-12-16
PRIOR APPLICATION NUMBER: JP 00/159162
PRIOR FILING DATE: 2000-04-07
PRIOR APPLICATION NUMBER: JP 00/280988
PRIOR FILING DATE: 2000-08-03
NUMBER OF SEQ ID NOS: 7059
SOFTWARE: PatentIn ver. 3.0
SEQ ID NO 4643
LENGTH: 325
TYPE: BPT
ORGANISM: Corynebacterium glutamicum
US-09-738-626-4643

```

Query Match	46.5%;	Score 755;	DB 10;	Length 325,
-------------	--------	------------	--------	-------------

Best Local Similarity 48.3%; Pred. NO. 2.5e-65;  
Matches 152; Conservative 67; Mismatches 86; Indels 10; Gaps 4;

QY	2	QIVLANPFCGVDRALEIVDOALKEFGAIIYRHVFNHRLTVDDQJCKRGAVFIELS	61
Db	15	KILMAPRGYCGVDRAVETVETRALKEEGARIVYRKELVHNRRVYVDLAEKGMIFVNEAS	74
QY	62	DVPVSYLIFSAGHSKEVQCEAEERKQLTVDPALCEPLVTKVHMQVAKHAAQGEVILGH	121
Db	75	EAPEGGANNWFSAHSGSPVYHBEAAAKNIKALDAACPLVTKHKEVORFDRQGHILFETGH	134
QY	122	AGHPVEGIMQYERKCTEGGIIYLVETPEDEVRLNK--VNNPNDLAVYIQTLLSWTDKVM	179
Db	135	BGHEEVBGMHNSVKEI-----HLVDGVAIGALTLPEFLNDEPNLIYMSQTLLSVTDMEI	189
QY	180	VDALNRQPSISIEQKKDILCYATQNRQAAVNDLAKISDILVYSGSPNNSNRRLREIAVQ	239
Db	190	VELKPKFPEQLDDPSDDICYATQNRQAAVNAIAERCELMIVGSRSSNSVRLVEVAQO	249
QY	240	LGRF--AYLIDTVQDDKQMLGIEIEVGVGTAGASAPVAVQEVIDQLKAMGGRTTSVRENS	298
Db	250	NGADNAAYLVADYARRIDPAMPFGEVETIGISSASVPBILVQGVIRLAEFGD--DVEEVT	307
QY	299	GIEEKVVPISPKELK	313
Db	308	SAAEKIVPALPRVLR	322

```

RESULT 6
US-09-712-363-271
Sequence 271, Application US/09712363
Patent No. US20020164588A1

GENERAL INFORMATION:
APPLICANT: Eisendberg, David
APPLICANT: Rotstein, Sergio H.
APPLICANT: Marcotte, Edward M.
TITLE OF INVENTION: DETERMINING THE FUNCTIONS AND
TITLE OR INVENTION: INTERACTIONS OF PROTEINS BY COMPARATIVE ANALYSIS
FILE REFERENCE: 07419-032001
CURRENT APPLICATION NUMBER: US/09/712,363
CURRENT FILING DATE: 2000-11-13
PRIOR APPLICATION NUMBER: PCT/US00/02246
PRIOR FILING DATE: 2000-01-28
PRIOR APPLICATION NUMBER: 60/179,531
PRIOR FILING DATE: 2000-02-01
PRIOR APPLICATION NUMBER: 60/117,844
PRIOR FILING DATE: 1999-01-29
PRIOR APPLICATION NUMBER: 60/118,206,
PRIOR FILING DATE: 1999-02-01
PRIOR APPLICATION NUMBER: 60/126,593
PRIOR FILING DATE: 1999-03-26
PRIOR APPLICATION NUMBER: 60/134,093
PRIOR FILING DATE: 1999-05-14
PRIOR APPLICATION NUMBER: 60/134,092
PRIOR FILING DATE: 1999-05-14
PRIOR APPLICATION NUMBER: 60/165,124
PRIOR FILING DATE: 1999-11-12
PRIOR APPLICATION NUMBER: 60/165,086
PRIOR FILING DATE: 1999-11-12
NUMBER OF SEQ ID NOS: 292
SOFTWARE: FastSeq for Windows Version 4.0
SEQ ID NO 271
LENGTH: 329
TYPE: prt
ORGANISM: Mycobacterium tuberculosis
US-09-712-363-271

```

```
Query Match      45.8%; Score 743.5; DB 10; Length 329;  
Best Local Similarity 47.68; Prid. No. 3,4e-64;  
Matches 150; Conservative 60; Mismatches 94; Indels 11; Gaps 3;  
  
3 IVLNPFGCGAGDRALEIVDAIEAFGAPITVRRHVNNRTVVNGLKQKGAVFIEELD 62  
::|||:|||:|||:|||::: |||::|::| |::|::|::|::|:
```

```
QY      3  IYLANPFGCAGUDRAIEIVDQALTEFGAPLYVRHEVHNRTVTDGLKKQGAVFIEELSD 62
```

Db 20 VELASPRSCAGVBAEIAETVKKVLDVAEGPVYVYKQIVANTVVAELDRGAVFEVDDLE 79  
 QY 63 V----PVSGYLIPSAHGVSKVQOBAERQLTVPATCPVTVMQVAKHAKQGREVLE 118  
 Db 80 IPDPPEPGAVVVSAGVSPAVFAGADERGLOVATCPVLAKVAKHAKFARQDITVFE 139  
 QY 119 IGHAGHPEVEGTMGQYEKCTEGGGIYLVETPEDVRNLKVNPNNDLAAYTQTLTSMTDKRV 178  
 Db 140 IGHAGHPEVEGTMGQYEKCTEGGGIYLVETPEDVRNLKVNPNNDLAAYTQTLTSMTDKRV 194  
 QY 179 MYDALREOPPSIKEQKDDICATONRODAVHDLAKISDLILVGSFNSNSNRLEIAV 238  
 Db 195 VIDALRAFPTLQGPSEDIQVATINRQALQSMWGECDVIVIGSCNNSNRRLVELAQ 254  
 QY 239 QUGKPAVILDTYQDLKQMLBGEIEVVGVTAGASAPVTVQEVLDLXKMGGETTSVRNS 298  
 Db 255 RSGTFAVILDGPDDILEPWLSSVSTIGVAGASAPRLVGVQVIALRGYASIT--VVERS 312  
 QY 299 GIEKRVFSIPKELX 313  
 Db 313 IATEVTRFGLPKQVR 327

## RESULT 7

US-09-712-363-188  
 ; Sequence 188, Application US/09712363  
 ; Patent No. US20020154588A1  
 ; GENERAL INFORMATION:  
 ; APPLICANT: Eisenberg, David  
 ; APPLICANT: Rotstein, Sergio H.  
 ; APPLICANT: Marcotte, Edward M.  
 ; TITLE OF INVENTION: DETERMINING THE FUNCTIONS AND  
 ; TITLE OF INVENTION: INTERACTIONS OF PROTEINS BY COMPARATIVE ANALYSIS  
 ; FILE REFERENCE: 07419-032091  
 ; CURRENT APPLICATION NUMBER: US/09/712,363  
 ; PRIOR FILING DATE: 2000-11-13  
 ; PRIOR APPLICATION NUMBER: PCT/US00/02246  
 ; PRIOR FILING DATE: 2000-01-28  
 ; PRIOR APPLICATION NUMBER: 60/179,531  
 ; PRIOR FILING DATE: 2000-02-01  
 ; PRIOR APPLICATION NUMBER: 60/117,844  
 ; PRIOR FILING DATE: 1999-01-29  
 ; PRIOR APPLICATION NUMBER: 60/118,206  
 ; PRIOR FILING DATE: 1999-02-01  
 ; PRIOR APPLICATION NUMBER: 60/126,593  
 ; PRIOR FILING DATE: 1999-03-26  
 ; PRIOR APPLICATION NUMBER: 60/134,093  
 ; PRIOR FILING DATE: 1999-05-14  
 ; PRIOR APPLICATION NUMBER: 60/134,092  
 ; PRIOR FILING DATE: 1999-05-14  
 ; PRIOR APPLICATION NUMBER: 60/165,124  
 ; PRIOR FILING DATE: 1999-11-12  
 ; PRIOR APPLICATION NUMBER: 60/165,086  
 ; PRIOR FILING DATE: 1999-11-12  
 ; NUMBER OF SEQ ID NOS: 292  
 ; SOFTWARE: FastSeq for Windows Version 4.0  
 ; SEQ ID NO 188  
 ; LENGTH: 335  
 ; TYPE: PRT  
 ; ORGANISM: Mycobacterium tuberculosis  
 US-09-712-363-188

Query Match 45.2%; Score 733; DB 10; Length 335;  
 Best Local Similarity 47.0%; Pred. No. 3,8e-63;  
 Matches 147; Conservative 66; Mismatches 92; Indels 8; Gaps 3;

QY 2 QIVLANPGRFCAGVDRAIEIVDOAIEAFGAPIYVRHEVNRITVVDGKQKGAVFIEBELS 61  
 Db 27 RVLLAEPGRGVCAGVDRAIEIVDOAIEAFGAPIYVRHEVNRITVVDGKQKGAVFIEBELS 86  
 QY 62 DVPVGSYLIPSAHGVSKVQOBAERQLTVPATCPVTVMQVAKHAKQGREVILIGH 121  
 Db 87 QVPEGAIVVFSAGVSPAVFAGADERGLOVATCPVLAKVAKHAKFARQDITVFE 146

QY 122 AGHPEVEGTMGQYEKCTEGGGIYLVETPEDVRNLKVNPNNDLAAYTQTLTSMTDKRV 181  
 Db 147 EGHEEVGTAGE-----APDVQVLDVGDVADVQVTVADEDEKVVMLTQTLTSMTDKRV 201  
 QY 182 ALREOPPSIKEQKDDICATONRODAVHDLAKISDLILVGSFNSNSNRLEIAV 241  
 Db 202 RLRRFFKLDQPPSEDIQVATINRQALQSMWGECDVIVIGSCNNSNRRLVELAQ 261  
 QY 242 KPAVILDTYQDLKQMLBGEIEVVGVTAGASAPVTVQEVLDLXKMGGETTSVRNSGI 300  
 Db 262 ARAHLVDMADIDISATLDGVTGVTAGASAPVTVQEVLDLXKMGGETTSVRNSGI 319  
 QY 301 EKVFSIPKELX 313  
 Db 320 NETLVFALPRELR 332

## RESULT 8

US-10-289-762-1090  
 ; Sequence 1090, Application US/10289762  
 ; Publication No. US20040006218A1  
 ; GENERAL INFORMATION:  
 ; APPLICANT: Griffiths, R.  
 ; TITLE OF INVENTION: Chlamydia pneumoniae genomic sequence and polypeptides, fragments  
 ; TITLE OF INVENTION: thereof and uses thereof, in particular for the diagnosis, prevention  
 ; FILE REFERENCE: 9710-003-999  
 ; CURRENT APPLICATION NUMBER: US/10/289,762  
 ; PRIOR FILING DATE: 2003-03-27  
 ; NUMBER OF SEQ ID NOS: 6849  
 ; SEQ ID NO 1090  
 ; TYPE: PRT  
 ; ORGANISM: Chlamydia pneumoniae  
 US-10-289-762-1090

Query Match 40.9%; Score 663; DB 12; Length 310;  
 Best Local Similarity 45.0%; Pred. No. 2.4e-56;  
 Matches 143; Conservative 59; Mismatches 98; Indels 18; Gaps 3;

QY 2 QIVLANPGRFCAGVDRAIEIVDOAIEAFGAPIYVRHEVNRITVVDGKQKGAVFIEBELS 61  
 Db 3 KLILCNERGPGCSGVAIVQVEVALERKMGAPIYKHEIVNRHVVALRAKGAIPVEBELV 62  
 QY 62 DVPVGSYLIPSAHGVSKVQOBAERQLTVPATCPVTVMQVAKHAKQGREVILIGH 121  
 Db 63 DVPVGSYLIPSAHGVSKVQOBAERQLTVPATCPVTVMQVAKHAKQGREVILIGH 122  
 QY 122 AGHPEVEGTMGQYEKCTEGGGIYLVETPEDVRNLKVNPNNDLAAYTQTLTSMTDKRV 181  
 Db 123 KKEVEVIGLVGEVPE-----HITVEKVDVLEALPSSSDPLFLYITQTLTSMTDKRV 177  
 QY 182 ALREOPPSIKEQKDDICATONRODAVHDLAKISDLILVGSFNSNSNRLEIAV 241  
 Db 178 ALKRIPISTLTPSSSICATINRQALQSMWGECDVIVIGSCNNSNRRLVELAQ 237  
 QY 242 KPAVILDTYQDLKQMLBGEIEVVGVTAGASAPVTVQEVLDLXKMGGETTSVRNSGI 301  
 Db 238 VPADLILNPPEDIDITNIVNSGDIAMTAGASTPEDVQACIRKL-----SLIPGQVE 290  
 QY 302 -----EKVFSIPKELX 313  
 Db 291 NDIFAVEDVVFOLPRELR 308

## RESULT 9

US-10-022-832-36  
 ; Sequence 36, Application US/10022832  
 ; Publication No. US20030059896A1  
 ; GENERAL INFORMATION:  
 ; APPLICANT: COUDRE, FRANCE  
 ; APPLICANT: HAMEL, JOSEF

Query Match	36.9%	Score 598.5;	DB 15;	Length 539;
Best Local Similarity	43.6%;	Pred. No. 1.2e-49;		
Matches 123; Conservative	55;	Mismatches 99;	Indels 5;	Gaps 1.

```

Query MatchScore 6.8%; Score 111; DB 15; Length 917;
Best local Similarity 20.7%; Pred. No. 0.14;
Matches 68; Conservative 50; Mismatches 97; Indels 114; Gaps 15

QY      4 VLAFRGFCAGDAIEIVD--QAIFAPAFIYVREHVNTVDGLKOKGAVFIEL 60
          ::||::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|
Db       632 IITNPKRAAEALQWVVREEMDLRYDDLAFG---YHIDDFNEAIRNG-KVK----- 678

QY      61 SDVPAGSYLISANGSVKEVQGEARESQLVPDATCELYTKNH-NQVAKHAQGSEVL 118
          :||::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|
Db       679 -LEPGS-----EKELSPPYLLVTVDELMDLMNVAPR----- 709

QY      119 IGHAHQPEVEGTMGQYEKCTEGGGIYLV-ETPE-DVRN-LKANPNPDLAVYTQTLS 172
          ::||::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|
Db       710 ----DVEDAIYRIQLARAGIHILVLTATGSPSDVVTGLIKANVPSRLAPTS--S 759

QY      173 MTDTKNWD-----ALREPFSSTKECK----- 194
          :|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|
Db       760 LADSRVILLDPQAEKLIGKGDELFLPKGANKPTRMGAFTEBEVEAVVQHCKDQAPVF 819

QY      195 KDDICAYATONQDAVHDIAKISDI-----LVYGSPSSNSNRILREIAVOLGPAVLDT 249
          ::||::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|
Db       820 RDDVTVGVKKQGIIDEIDIGDDLDLLCGAAELVISTFGSTSMQRCLRGVGPAKAGIMDL 879

QY      250 YQDLKQDMLGIEVVGVAAGAASPEVIVQ 278
          :|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|

```

DB 880 -----MESNRIVSGSESKARDLVNK 900

## RESULT 12

US-09-712-363-282  
 / Sequence 282, Application US/09712363  
 / Patent No. US20020164588A1  
 / GENERAL INFORMATION:  
 / APPLICANT: Eisenberg, David  
 / APPLICANT: Rotstein, Sergio H.  
 / APPLICANT: Marcotte, Edward M.  
 / TITLE OF INVENTION: DETERMINING THE FUNCTIONS AND  
 / TITLE OF INVENTION: INTERACTIONS OF PROTEINS BY COMPARATIVE ANALYSIS  
 / FILE REFERENCE: 07419-032001  
 / CURRENT APPLICATION NUMBER: US/09/712,363  
 / PRIOR FILING DATE: 2000-11-13  
 / PRIOR APPLICATION NUMBER: PCT/US00/02246  
 / PRIOR FILING DATE: 2000-01-28  
 / PRIOR APPLICATION NUMBER: 60/179,531  
 / PRIOR FILING DATE: 2000-02-01  
 / PRIOR APPLICATION NUMBER: 60/117,844  
 / PRIOR FILING DATE: 1999-01-29  
 / PRIOR APPLICATION NUMBER: 60/118,206  
 / PRIOR FILING DATE: 1999-02-01  
 / PRIOR APPLICATION NUMBER: 60/126,593  
 / PRIOR FILING DATE: 1999-03-26  
 / PRIOR APPLICATION NUMBER: 60/134,093  
 / PRIOR FILING DATE: 1999-05-14  
 / PRIOR APPLICATION NUMBER: 60/134,092  
 / PRIOR FILING DATE: 1999-05-14  
 / PRIOR APPLICATION NUMBER: 60/165,124  
 / PRIOR FILING DATE: 1999-11-12  
 / PRIOR APPLICATION NUMBER: 60/165,086  
 / PRIOR FILING DATE: 1999-11-12  
 / NUMBER OF SEQ ID NOS: 292  
 / SOFTWARE: FastSeq for Windows Version 4.0  
 / SEQ ID NO: 282  
 / LENGTH: 578  
 / TYPE: PRT  
 / ORGANISM: Mycobacterium tuberculosis  
 / US-09-712-363-282

Query Match 6.7%; Score 109; DB 10; Length 578;  
 Best Local Similarity 20.4%; Pred. No. 0.11;  
 Matches 78; Conservative 61; Mismatches 105; Indels 138; Gaps 18;

QY 6 ANPRGFC-----AGVRAIEVDQALZAFGAPLYVEEYVH 41  
 DB 67 ANPCGVCESVSLAPAPGSDIVVELDASHGVDTRDLDR--AFYAPQSRVY-- 121  
 QY 42 NRTVVDGLKQKAVFIEELSDVPVGSYLLFFSAHGVSK-----EVQGEAEERQTLVDA 94  
 DB 122 -----FIVDEHMYVTYTGFMNLLKIVEPPEHLIFITAT 155  
 QY 95 TCP-----LVTKVH-----MQVAKAKOGEREVILIGHAGHPEVEGTMCQYE 135  
 DB 156 TEPEKVLPIRSRTHYPRFLPPRTMLARICEQ--EGVVDVAVPLV----- 205  
 QY 136 KCTBGGGILVETPEB--VRMLKANNPVDLAYVTQT-----TLSTDTKAMVDALREQPP 188  
 DB 206 -IRAGGG-----SPRDTLSVLDQLLAGAAD-THVYTRALGLLGYTVALIDDAV----D 254  
 QY 189 SIKGCKDDICATVGNRODAVHD-----LAKISDLIVVGSFNSNS----- 230  
 DB 255 ALAACDAALFGLAISVIDIGGHPRRFATDLERRODLIVOSVDAASRGVDAPEBDAL 314  
 QY 231 NLRLEIAVLGKPAVLIIDTYQDLKQDMLEGIIVGVTAGASAPVLYQCEVIDQL--KAM 287  
 DB 315 DMKREQAARIGAT--LRTIYAVVQAGL-----GEMRGATAPRLLLEVVCARLLLPAS 366  
 QY 288 GGETTSVRENSGIEEKVPSIP 309  
 DB 367 DAEASALGVEVERIEIRLDMKIP 388

## RESULT 13

US-10-369-493-5415  
 / Sequence 5415, Application US/10369493  
 / Publication No. US20030233675A1  
 / GENERAL INFORMATION:  
 / APPLICANT: Cao, Yongwei  
 / APPLICANT: Hinkle, Gregory J.  
 / APPLICANT: Slater, Steven C.  
 / APPLICANT: Goldman, Barry S.  
 / APPLICANT: Chen, Xianfeng  
 / TITLE OF INVENTION: EXPRESSION OF MICROBIAL PROTEINS IN PLANTS FOR PRODUCTION OF  
 / TITLE OF INVENTION: PLANTS WITH IMPROVED PROPERTIES  
 / FILE REFERENCE: 38-10152052/B  
 / CURRENT APPLICATION NUMBER: US/10/369,493  
 / PRIOR FILING DATE: 2003-02-28  
 / PRIOR APPLICATION NUMBER: US 60/360,039  
 / PRIOR FILING DATE: 2002-02-21  
 / NUMBER OF SEQ ID NOS: 47374  
 / SEQ ID NO: 5415  
 / LENGTH: 925  
 / TYPE: PRT  
 / ORGANISM: Caenorhabditis elegans  
 / US-10-369-493-5415

Query Match 6.7%; Score 109; DB 12; Length 925;  
 Best Local Similarity 23.1%; Pred. No. 0.23;  
 Matches 77; Conservative 53; Mismatches 134; Indels 70; Gaps 17;

QY 3 IVLANPRGFCAGVRAIEVDQALZAFGAPLYVEEYVHNTVVDGLKQKAVFIEELSD 62  
 DB 143 MILNARKFTIEEYVQAAEIKQVVT--VPYF--TPAEELAVRAAQVAGLTVLDLIN 196  
 QY 63 VPVGSYLLFFSAHGVSKVQGEAEERQTLVDA--ACPLVTYKVMQVAKAKOGEREVIL 118  
 DB 197 DGTAAAL---SHGIFRKEIGEKNQRLMYDMGAKTATIVERKLVEKERYQPKYTVL 253  
 QY 119 -IGHAGHPEVEGTMCQYEKCTBGGGILVETPEDEVRLKVN--NPPVDLAYVTQTLSMT 174  
 DB 254 GVG-----PDRITGGIE-WTNRRLRHLIEMFE--KYYKRTKYNTVTRRA-MTYSKEAR 303  
 QY 175 DTKVVDALREQFISIEQKDDICVATNORODAVHDLAKISDLIVVGSFNSNSRLR 234  
 DB 304 RLKQVLANNEHFPQISAHEDIDAKLVTREDYH--LISWESRFGEI-IGQALRMA 359  
 QY 235 EIAVQDLKPAVLIIDTYQDLKQDMLEGIIVGVTAGASAPV--LVQSEV----- 281  
 DB 360 QLP-----DDIDQFVLMG-----AGTRVPKQVEIVQKTIQKXIGKFLMT 400  
 QY 282 DQKAMGGETTSVRENSG-----IEEKVPSI 308  
 DB 401 DEAVAMGALFQAAHLSKGFYKPPNIEKQVLPV 434

## RESULT 14

US-10-369-493-1052  
 / Sequence 1052, Application US/10369493  
 / Publication No. US20030233675A1  
 / GENERAL INFORMATION:  
 / APPLICANT: Cao, Yongwei  
 / APPLICANT: Hinkle, Gregory J.  
 / APPLICANT: Slater, Steven C.  
 / APPLICANT: Goldman, Barry S.  
 / APPLICANT: Chen, Xianfeng  
 / TITLE OF INVENTION: EXPRESSION OF MICROBIAL PROTEINS IN PLANTS FOR PRODUCTION OF  
 / TITLE OF INVENTION: PLANTS WITH IMPROVED PROPERTIES  
 / FILE REFERENCE: 38-10152052/B  
 / CURRENT APPLICATION NUMBER: US/10/369,493  
 / PRIOR FILING DATE: 2003-02-28  
 / PRIOR APPLICATION NUMBER: US 60/360,039  
 / PRIOR FILING DATE: 2002-02-21  
 / NUMBER OF SEQ ID NOS: 47374

SEQ ID NO 1052  
LENGTH: 501  
TYPE: PRT  
ORGANISM: Methanococcus jannaschii  
US-10-369-493-1052

Query Match  
Best Local Similarity 20.5%; Pred. No. 0.24;  
Matches 69; Conservative 51; Mismatches 134; Indels 83; Gaps 15;

14 GVDRAIEIVDAI--EAFGAPIYR-----HEVVH-RTVVDGLK 50  
133 GSEGAIEDIDEAIEIAEAGFPVVVYASAGGGMGVAVSKEKLEVESARINAKAF 192  
51 QGKAVFIEELSDVP--VGSYLIFSAHG-----VSKVQGAHEERQLTVDACPLVTKVH 103  
193 GDFPIETIEKLEPNPHIEIQLGDRNGNIIHJADRCSTIQSHQKI--LEEAPSPMTB-- 249  
104 MGVAKAKOGREVLIGHAGHEPEVGTMGQVEKTEGGGIVLVETPEDEVRNKKVNPDL 163  
250 ---ELRRWGEAIAIKGKAINDSAGTV---EFLVENGFIYLE-----MNRIGVEHT- 297  
164 AVYQTLTSMITKYKWDALREQPSIKEQKDDICVATQ--NFQAVHDIAKISDLIL 220  
298 --VTEQVTGIDLVKMIKILAGEBELTKQEDVKIRGHAIKCRINADPLNDPVPKPGKIK 355  
221 VVGSPPNSNSNRLRIAVOLGKPAVLIPTYDPLKQDWLEGIWGVVGTAGASAPVLAQEV 280  
356 LYRSP-----GGPVRIIDB-----GVYGAELIP--YDSM 384  
281 IDQKAMG--ETTSVRENSGIEEKVPSIPKELKH 315  
385 IAKLITIGNSREEAIAKMKRALREYVIGVKNIPFH 421

RESULT 15

US-10-369-493-19263  
Sequence 19263, Application US/10369493  
Publication No. US2003023675A1  
GENERAL INFORMATION:  
APPLICANT: Cao, Yongwei  
APPLICANT: Hinkle, Gregory J.  
APPLICANT: Slater, Steven C.  
APPLICANT: Goldman, Barry S.  
APPLICANT: Chen, Xiandeng  
TITLE OF INVENTION: EXPRESSION OF MICROBIAL PROTEINS IN PLANTS FOR PRODUCTION OF  
FILE REFERENCE: 38-10(52052)B  
CURRENT APPLICATION NUMBER: US/10/369,493  
PRIOR FILING DATE: 2003-02-28  
PRIOR APPLICATION NUMBER: US 60/360,039  
NUMBER OF SEQ ID NOS: 47374  
SEQ ID NO 19263  
LENGTH: 1130  
TYPE: PRT  
ORGANISM: Myxococcus xanthus  
US-10-369-493-19263

Query Match  
Best Local Similarity 6.4%; Score 104.5; DB 12; Length 1130;  
Matches 70; Conservative 64; Mismatches 124; Indels 155; Gaps 13;

38 EYVNRRTVVDGLKQKAVFIEELSDV-----PVGSYLI-----FSA 73  
604 EVAEIADINAMTQTLISIFAQOVTDVARTVGEKLAGAQAAPVAGATWKDILINNVLAA 663  
74 HGVSKVQGAHEERQLTVDACPLVTKVH 104  
664 NNLTIDGVNINAEVTTAVAKDLSKRTIVDAKGEVLELSTINTMVDQLNSPAAEVTRAK 723  
105 QVAKAKOGREVLIGHAGHEPEVGTMGQVEKTEGGGIVLVETPEDEV 153

DB 724 EVGTGKLG-----QGAPEVGVSWKMDLIDNNFMAVNLITQVRIYRVTAVAN 774  
QY 154 ---NLKVNPNDLAVYTQTLTSMITD---KQWDAIR-----EQFPSIKE 192  
DB 775 GDINQKLTMDAKGEIAALADITINAMTQTLISIFAQOVTDVARTVGEKLAGAQAEPVAVAG 834  
QY 193 QKXD---DICATONRQDAVHDIKIS-----DL-----LIVGSPPNSNSNR 232  
DB 835 TWKDLTNVNNILANNLTIAQVRINAEVTTAVANGDLSKRTIVDAKGEVLELSTINTMVDQ 894  
QY 233 LREIAVOLGKPAVLIIDT-----YQDLKQDWLEGIWGVVGTAGASAPV----- 275  
DB 895 LRAFAAVTRVAKGEVGTGKLGQGAADVYDLGKWKDLIDNNVNLAGNLTQVRNIAKTYT 954  
QY 276 -----LVQEVTDQKAMGSETTSVRENSGIEKV 304  
DB 955 AVANGDLSQKITVSVKGEVLELKNITINTMVDQLAFASEVTRVAKGEVGTGKLG 1007

Search completed: January 29, 2004, 16:21:19  
Job time : 22.7073 secs

GenCore version 5.1.6  
Copyright (c) 1993 - 2004 Compugen Ltd.

OM protein - protein search, using sw model

Run on: January 29, 2004, 15:44:49 ; Search time 9.35661 Seconds  
(without alignments)  
3268.453 Million cell updates/sec

Title: US-09-941-947a-18

Perfect score: 1623

Sequence: 1 MGIYANPRGFCAGVDRAIE.....GIEKVFSPKELKMGQA 318

Scoring table: BLOSUM62  
Gapop 10.0 , Gapext 0.5

Searched: 283308 seqs, 96168682 residues

Total number of hits satisfying chosen parameters: 283308

Minimum DB seq length: 0

Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

1: PIR.\*  
2: PIR.\*  
3: PIR.\*  
4: PIR.\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

# SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	1028.5	63.4	314	2	G83076
2	1019.5	62.8	317	2	AC0059
3	1001	61.7	316	2	AD0508
4	996	61.4	316	2	JB0403
5	996	61.4	316	2	H90632
6	987	60.8	320	2	G82293
7	987	60.8	320	2	G82293
8	969	59.7	318	1	A64164
9	968	59.6	322	2	D81038
10	967	59.6	316	2	C82561
11	858	52.9	319	2	A84947
12	826.5	50.9	319	2	G87665
13	756	46.6	348	2	AH2671
14	756	46.6	348	2	F97453
15	750	46.2	346	2	AE3434
16	743.5	45.8	329	2	E70973
17	733	45.2	335	2	D70898
18	717	44.2	335	2	E87151
19	667	41.1	307	2	B71461
20	663	40.9	310	2	G72006
21	663	40.9	310	2	F86617
22	661	40.7	308	2	C81723
23	621.5	38.3	180	2	D37152
24	565	34.8	335	2	G75939
25	527	32.5	314	1	B69554
26	513	31.6	314	2	F83822
27	443.5	27.3	311	2	AC1256
28	415.5	25.6	289	2	G70449
29	369.5	22.8	277	2	F81362

30	358	22.1	642	2	H97127	fusion Penicillin
31	341	21.0	446	2	S76307	hypothetical prote
32	335.5	20.7	274	1	H64569	lysis tolerance pr
33	332	20.5	275	2	A72253	lytB protein - The
34	312	19.2	402	2	AF1929	hypothetical prote
35	288	17.7	376	2	G71310	probable penicilli
36	230	14.2	452	2	T04781	hypothetical prote
37	115	7.1	929	2	T35683	ftsk homolog - Str
38	112.5	6.9	434	2	B72228	pyrimidine-nucleos
39	111.5	6.9	777	1	G69773	conserved hypotet
40	110	6.8	500	2	AB1111	heat-shock protein
41	109	6.7	499	2	A11472	heat-shock protein
42	109	6.7	522	2	A33644	signal recognition
43	109	6.7	578	2	B70796	probable DNA POLYX
44	109	6.7	925	2	G88175	protein T247.2 [l
45	108.5	6.7	544	2	B41479	60K heat shock pro

## ALIGNMENTS

### RESULT 1

G83076 lytB protein PA4557 [imported] - Pseudomonas aeruginosa (strain PA01)

C/Species: Pseudomonas aeruginosa

C/Date: 15-Sep-2000 #sequence\_revision 15-Sep-2000 #ext\_change 31-Dec-2000

C/Accession: G83076

R/Stover, C.R.; Pham, X.O.; Erwin, A.L.; Mizoguchi, S.D.; Warren, P.; Hickey, M.J.; B

adman, S.; Yuan, Y.; Brody, L.L.; Coulter, S.N.; Folger, K.R.; Kas, A.; Larbig, K.; Lin

., Lory, S.; Olson, M.V.

Nature 406, 959-964, 2000

A/Title: Complete genome sequence of Pseudomonas aeruginosa PA01, an opportunistic path

A/Reference number: A82950; PMID:20437337; PMID:10984043

A/Accession: G83076

A/Status: preliminary

A/Molecule type: DNA

A/Restrictions: 1-314 <STO>

A/Cross-references: GB:AE004869; GB:AE004091; NID:g9950793; PIDN:AA07945.1; GSPDB:GN00

A/Experimental source: strain PA01

C/Genetics:

A/Gene: lytB; PA4557

C/Superfamily: penicillin tolerance protein

Query Match 63.4%; Score 1028.5; DB 2; Length 314;

Best Local Similarity 63.3%; Pred. No. 6.8e-67;

Matches 198; Conservative 49; Mismatches 63; Indels 3; Gaps 2;

QY	1	MGIYANPRGFCAGVDRAIEITVDQALFAFGAIIYRHEVYVNRRTYVDGLKQKGAVFIEEL	60
DB	1	MGIYANPRGFCAGVDRAIEITVDQALFAFGAIIYRHEVYVNRRTYVDGLKQKGAVFIEEL	60
QY	61	SDVPVGSYLIFSAHGVSKVEQDEAEEROLTVDATCPVATKVMQVAKAKQKQREYILIG	120
DB	61	SDVPVNNVITVIFSAHGVSKVEQDEAEEROLTVDATCPVATKVMQVAKAKQKQREYILIG	120
QY	121	HAGHPVEBGTGQYKCTEGGSIYVETPEDVRNKKVNNPNDAVYVTTLSMTDTKVMV	180
DB	121	HAGHPVEBGTGQYKCTEGGSIYVETPEDVRNKKVNNPNDAVYVTTLSMTDTKVMV	180
QY	181	DALREOPSIKQKQDDICVANTONRQDAVHDIAKISDILIVAGSSNSNSNRLREIAVOL	240
DB	181	DALREOPSIKQKQDDICVANTONRQDAVHDIAKISDILIVAGSSNSNSNRLREIAVOL	240
QY	241	GXPAYLIDTYODLKQDMLEGIEVGVTAQASPEVLVQEVVIDOLKAMGGETTSVRENSGI	300
DB	241	GXPAYLIDTYODLKQDMLEGIEVGVTAQASPEVLVQEVVIDOLKAMGGETTSVRENSGI	300
QY	301	EKGVFSPKELK 313	
DB	301	EKGVFSPKELK 313	
QY	298	ENITFSMKELK 310	
DB	298	ENITFSMKELK 310	

RESULT 2

AC0058

lytB protein [imported] - Yersinia pestis (strain CO92)

C/Species: Yersinia pestis

C/Date: 02-Nov-2001 #sequence\_revision 02-Nov-2001 #text\_change 09-Nov-2001

C/Accession: AC0058

R/Parikh, J.; Wren, B.W.; Thomson, N.R.; Tilball, R.W.; Holden, M.T.G.; Prentice, M.B.;

deno-Tarrega, A.M.; Chillingworth, T.; Cronin, A.; Davies, R.M.; Davis, P.; Dougan, G.;

11, M.; Rutherford, K.; Simmonds, M.; Skelton, J.; Stevens, K.; Whitehead, S.; Barrett,

Nature 413, 523-527, 2001

A/Title: Genome sequence of Yersinia pestis, the causative agent of plague.

A/Reference number: AB0001; MUID:21470413; PMID:11586360

A/Accession: AC0059

A/Status: Preliminary

A/Molecule type: DNA

A/Residues: 1-317 &lt;YUR&gt;

A/Cross-References: GB:AL530842; PIDN:CA089334.1; PID:g15978571; GSPDB:GN00175

C/Genetics:

A:Gene: lytB

C:Superfamily: penicillin tolerance protein

Query Match 62.8%; Score 1019.5; DB 2; Length 317;

Best Local Similarity 63.5%; Pred. No. 3.1e-66;

Matches 200; Conservative 54; Mismatches 54; Indels 7; Gaps 4;

QY 1 MOVLNPRGFCAGVDRAIEIVDAIEAFGAPIVRHEVHNRTVVDLKKQGAFTIEL 60

DB 1 MOVLNPRGFCAGVDRAIEIVDAIEAFGAPIVRHEVHNRTVVDLKKQGAFTIEL 60

QY 61 SDVPVGSYLFSAHGVSKVQQAEEERQLT-VPDAPLVTYKVMQVAKAAGREVILI 119

DB 61 SEVPDGSILIFSAHGVSKVQQAEEERQLT-VPDAPLVTYKVMQVAKAAGREVILI 120

QY 120 GHAGPEVETMGVYKCTEGCGIYVETPEDVANKVNPNDLAVYTOTTLSMTDTKVM 179

DB 120 GHAGPEVETMGVYKCTEGCGIYVETPEDVANKVNPNDLAVYTOTTLSMTDTKVM 179

QY 121 GHAGPEVETMGVYKCTEGCGIYVETPEDVANKVNPNDLAVYTOTTLSMTDTKVM 178

DB 121 GHAGPEVETMGVYKCTEGCGIYVETPEDVANKVNPNDLAVYTOTTLSMTDTKVM 178

QY 180 VDAPPEPSTIKKQKODICATONRQDAVHDLAKISDLILVGSPPNSNSRLREIAVOL 239

DB 180 VDAPPEPSTIKKQKODICATONRQDAVHDLAKISDLILVGSPPNSNSRLREIAVOL 239

QY 179 IDANTREPKIVGPRKODICATONRQDAVHDLAKISDLILVGSPPNSNSRLREIAVOL 238

DB 179 IDANTREPKIVGPRKODICATONRQDAVHDLAKISDLILVGSPPNSNSRLREIAVOL 238

QY 240 LGKPAVILDTYQDKQKODICATONRQDAVHDLAKISDLILVGSPPNSNSRLREIAVOL 298

DB 240 LGKPAVILDTYQDKQKODICATONRQDAVHDLAKISDLILVGSPPNSNSRLREIAVOL 298

QY 239 MGKPAVILDSAADIQEFNLQAKICITVTAASAPDILVQVIRLDDLAGESIT---EIS 295

DB 239 MGKPAVILDSAADIQEFNLQAKICITVTAASAPDILVQVIRLDDLAGESIT---EIS 295

QY 299 GIEKVESIPREIK 313

DB 299 GIEKVESIPREIK 313

QY 236 GREENIVEVPEKELR 310

DB 236 GREENIVEVPEKELR 310

RESULT 3

lytB protein [imported] - Salmonella enterica subsp. enterica serovar Typh (strain CT18

C/Species: Salmonella enterica subsp. enterica serovar Typh

C/Date: 09-Nov-2001 #sequence\_revision 09-Nov-2001 #text\_change 18-Nov-2002

R/Parikh, J.; Dougan, G.; James, K.D.; Thomson, N.R.; Pickard, D.; Wain, J.; Churcher,

th, T.; Comerford, P.; Cronin, A.; Davies, P.; Davies, R.M.; Dowd, L.; White, N.; Farrar,

S.; Mouton, S.; O'Garra, P.

Nature 413, 848-852, 2001

A/Title: Complete genome sequence of a multiple drug resistant Salmonella enterica serov

A/Reference number: AB0502; MUID:21534947; PMID:11677608

A/Accession: AC0508

A/Status: Preliminary

A/Molecule type: DNA

A/Residues: 1-316 &lt;PAR&gt;

A/Cross-References: GB:AL513382; PIDN:CAD01204.1; PID:g16501333; GSPDB:GN00176

C/Genetics:

A:Gene: STY0058

C:Superfamily: penicillin tolerance protein

Query Match 61.7%; Score 1001; DB 2; Length 316;

Best Local Similarity 62.3%; Pred. No. 6.7e-65;

Matches 195; Conservative 53; Mismatches 61; Indels 4; Gaps 2;

QY 1 MOVLNPRGFCAGVDRAIEIVDAIEAFGAPIVRHEVHNRTVVDLKKQGAFTIEL 60

DB 1 MOVLNPRGFCAGVDRAIEIVDAIEAFGAPIVRHEVHNRTVVDLKKQGAFTIEL 60

QY 61 SDVPVGSYLFSAHGVSKVQQAEEERQLT-VPDAPLVTYKVMQVAKAAGREVILI 120

DB 61 SDVPVGSYLFSAHGVSKVQQAEEERQLT-VPDAPLVTYKVMQVAKAAGREVILI 120

QY 121 GHAGPEVETMGVYKCTEGCGIYVETPEDVANKVNPNDLAVYTOTTLSMTDTKVM 180

DB 121 GHAGPEVETMGVYKCTEGCGIYVETPEDVANKVNPNDLAVYTOTTLSMTDTKVM 180

QY 121 GHAGPEVETMGVYKCTEGCGIYVETPEDVANKVNPNDLAVYTOTTLSMTDTKVM 178

DB 121 GHAGPEVETMGVYKCTEGCGIYVETPEDVANKVNPNDLAVYTOTTLSMTDTKVM 178

QY 181 DALREPPSTIKKQKODICATONRQDAVHDLAKISDLILVGSPPNSNSRLREIAVOL 240

DB 181 DALREPPSTIKKQKODICATONRQDAVHDLAKISDLILVGSPPNSNSRLREIAVOL 240

QY 179 DALREPPSTIKKQKODICATONRQDAVHDLAKISDLILVGSPPNSNSRLREIAVOL 238

DB 179 DALREPPSTIKKQKODICATONRQDAVHDLAKISDLILVGSPPNSNSRLREIAVOL 238

QY 241 GKPAVILDTYQDKQKODICATONRQDAVHDLAKISDLILVGSPPNSNSRLREIAVOL 300

DB 241 GKPAVILDTYQDKQKODICATONRQDAVHDLAKISDLILVGSPPNSNSRLREIAVOL 300

QY 239 GRKPAVILDSAADIQEFNLQAKICITVTAASAPDILVQVIRLDDLAGESIT---GR 296

DB 239 GRKPAVILDSAADIQEFNLQAKICITVTAASAPDILVQVIRLDDLAGESIT---GR 296

QY 301 EEKVESIPREIK 313

DB 301 EEKVESIPREIK 313

QY 297 EENIVEVPEKELR 309

DB 297 EENIVEVPEKELR 309

RESULT 4

lytB protein - Escherichia coli (strain K-12)

C/Species: Escherichia coli

C/Date: 12-Mar-1993 #sequence\_revision 12-Mar-1993 #text\_change 01-Mar-2002

C/Accession: JE0403; S40552; E64723; S22290

R/Bowyer, J.; Stragier, P.

Nucleic Acids Res. 19, 180, 1991

A/Title: Nucleotide sequence of the *isp-dabB* interval in Escherichia coli.

A/Reference number: JE0403; MUID:91187617; PMID:2011499

A/Accession: JE0403

A/Molecule type: DNA

A/Residues: 1-316 &lt;YUR&gt;

A/Cross-References: EMBL:X54925

A/Reference number: JE0403; MUID:91187617; PMID:2011499

A/Accession: JE0403

A/Status: Preliminary

A/Molecule type: DNA

A/Residues: 1-316 &lt;YUR&gt;

A/Cross-References: EMBL:D10483; NID:9216434; PIDN:BA01307.1; PID:g216456

A/Experimental source: strain K-12, substrain MG1655

C/Genetics:

A:Gene: lytB

C:Superfamily: penicillin tolerance protein

Query Match 61.4%; Score 996; DB 2; Length 316;

Best Local Similarity 61.3%; Pred. No. 1.5e-64;

Matches 192; Conservative 59; Mismatches 58; Indels 4; Gaps 2;

QY 1 MOVLNPRGFCAGVDRAIEIVDAIEAFGAPIVRHEVHNRTVVDLKKQGAFTIEL 60

DB 1 MOVLNPRGFCAGVDRAIEIVDAIEAFGAPIVRHEVHNRTVVDLKKQGAFTIEL 60

QY 61 SDVPVGSYLIFSAHGVSKVQQAEROLTVFADATCPITVKNQVAKHAKQREVLIG 120  
 Db 61 SEVPDGMILIFSAHGVSKVQQAEROLTVFADATCPITVKNQVAKHAKQREVLIG 120  
 QY 121 HAGHPEVGTMGQYKCTEGGGITLVETPEDVRLKVNNDLAVYQTTLSTMTDKVMV 180  
 Db 121 HAGHPEVGTMGQYKCTEGGGITLVETPEDVRLKVNNDLAVYQTTLSTMTDKVMV 180  
 QY 181 DALREOPFSIKKQKDDICVATONRODAVHDLAKISDLILVGS PNSSNRLAEIAYOL 240  
 Db 179 DALRKRPKIVGPKKDDICVATONRODAVHDLAKISDLILVGS PNSSNRLAEIAYOL 238  
 QY 241 GKRAVILDTYODLKQMLBEGIEVVGTAAGASAPVILQVLDLKAQGETTSVRENSGI 300  
 Db 239 GKRAFLIDDAKDIOBEMVKEVKCVGTAGASAPDILVQNVARLQQLGGGEALPLE--GR 296  
 QY 301 EEKAVFSIPKELK 313  
 Db 297 EENIVEFVPEKLR 309

## RESULT 5

H90632

LybB protein [imported] - Escherichia coli (strain O157:H7, substrain RMD 0509952)  
 C/Species: Escherichia coli  
 C/Date: 18-Jul-2001 #sequence\_revision 18-Jul-2001 #text\_change 03-Aug-2001  
 C/Accession: H90632  
 R/Hayashi, T.; Makino, K.; Ohnishi, M.; Kurokawa, K.; Ishii, K.; Yokoyama, K.; Han, C.G.  
 Gasawara, N.; Yasunaga, T.; Kihara, S.; Shiba, T.; Hattori, M.; Shinagawa, H.  
 DNA Res. 9, 11-22, 2001  
 A/Title: Complete genome sequence of enterohemorrhagic Escherichia coli O157:H7 and genod  
 A/Reference number: A96629; MUID:21156231; PMID:11258796  
 A/Accession: H90632  
 A/Status: preliminary  
 A/Molecule type: DNA  
 A/Residues: 1-316 <HAV>  
 A/Cross-references: GB:BA000007; PTDN:BAE33455.1; PTD:G13359488; GSPDB:GN00154  
 A/Experimental source: strain O157:H7, substrain RMD 0509952  
 C/Genetics:  
 A/Gene: EC90032  
 C/Superfamily: penicillin tolerance protein

Query Match 61.4%; Score 996; DB 2; Length 316;  
 Best Local Similarity 61.3%; Pred. No. 1.5e-64;  
 Matches 192; Conservative 59; Mismatches 58; Indels 4; Gaps 2;  
 QY 1 MQVLANPRGFCAGVDRAIEIVDAIEAFGAPIVRHEVHNRTVVDGKQKAVFIEEL 60  
 Db 1 MQLILANPRGFCAGVDRAIEIVDAIEAFGAPIVRHEVHNRTVVDGKQKAVFIEEL 60  
 QY 61 SDVPVGSYLIFSAHGVSKVQQAEROLTVFADATCPITVKNQVAKHAKQREVLIG 120  
 Db 61 SEVPDGMILIFSAHGVSKVQQAEROLTVFADATCPITVKNQVAKHAKQREVLIG 120  
 QY 121 HAGHPEVGTMGQYKCTEGGGITLVETPEDVRLKVNNDLAVYQTTLSTMTDKVMV 180  
 Db 121 HAGHPEVGTMGQYKCTEGGGITLVETPEDVRLKVNNDLAVYQTTLSTMTDKVMV 180  
 QY 181 DALREOPFSIKKQKDDICVATONRODAVHDLAKISDLILVGS PNSSNRLAEIAYOL 240  
 Db 179 DALRKRPKIVGPKKDDICVATONRODAVHDLAKISDLILVGS PNSSNRLAEIAYOL 238  
 QY 241 GKRAVILDTYODLKQMLBEGIEVVGTAAGASAPVILQVLDLKAQGETTSVRENSGI 300  
 Db 239 GKRAFLIDDAKDIOBEMVKEVKCVGTAGASAPDILVQNVARLQQLGGGEALPLE--GR 296  
 QY 301 EEKAVFSIPKELK 313  
 Db 297 EENIVEFVPEKLR 309

## RESULT 6

85483  
 hypothetical protein lybB [imported] - Escherichia coli (strain O157:H7, substrain EDL93)  
 C/Species: Escherichia coli  
 C/Date: 16-Feb-2001 #sequence\_revision 16-Feb-2001 #text\_change 14-Sep-2001  
 C/Accession: G85483  
 R/Perna, N.T.; Plunkett III, G.; Burland, V.; Mau, B.; Glasner, J.D.; Rose, D.J.; Mayhew,  
 Miller, L.; Grochbeck, C.J.; Davis, N.W.; Lim, A.; Dimalanta, E.; Potamousie, K.; Apodaca,  
 Nature 409, 529-533, 2001  
 A/Title: Genome sequence of enterohemorrhagic Escherichia coli O157:H7.  
 A/Reference number: A85480; MUID:21074935; PMID:11206551  
 A/Accession: G85483  
 A/Status: preliminary  
 A/Molecule type: DNA  
 A/Residues: 1-316 <STO>  
 A/Cross-references: GB:AE005174; NID:G12512712; PTDN:AA654331.1; GSPDB:GN00145; UKGP:200  
 A/Experimental source: strain O157:H7, substrain EDL93  
 C/Genetics:  
 A/Gene: lybB  
 C/Superfamily: penicillin tolerance protein

Query Match 61.4%; Score 996; DB 2; Length 316;  
 Best Local Similarity 61.3%; Pred. No. 1.5e-64;  
 Matches 192; Conservative 59; Mismatches 58; Indels 4; Gaps 2;  
 QY 1 MQVLANPRGFCAGVDRAIEIVDAIEAFGAPIVRHEVHNRTVVDGKQKAVFIEEL 60  
 Db 1 MQLILANPRGFCAGVDRAIEIVDAIEAFGAPIVRHEVHNRTVVDGKQKAVFIEEL 60  
 QY 61 SDVPVGSYLIFSAHGVSKVQQAEROLTVFADATCPITVKNQVAKHAKQREVLIG 120  
 Db 61 SEVPDGMILIFSAHGVSKVQQAEROLTVFADATCPITVKNQVAKHAKQREVLIG 120  
 QY 121 HAGHPEVGTMGQYKCTEGGGITLVETPEDVRLKVNNDLAVYQTTLSTMTDKVMV 180  
 Db 121 HAGHPEVGTMGQYKCTEGGGITLVETPEDVRLKVNNDLAVYQTTLSTMTDKVMV 180  
 QY 181 DALREOPFSIKKQKDDICVATONRODAVHDLAKISDLILVGS PNSSNRLAEIAYOL 240  
 Db 179 DALRKRPKIVGPKKDDICVATONRODAVHDLAKISDLILVGS PNSSNRLAEIAYOL 238  
 QY 241 GKRAVILDTYODLKQMLBEGIEVVGTAAGASAPVILQVLDLKAQGETTSVRENSGI 300  
 Db 239 GKRAFLIDDAKDIOBEMVKEVKCVGTAGASAPDILVQNVARLQQLGGGEALPLE--GR 296  
 QY 301 EEKAVFSIPKELK 313  
 Db 297 EENIVEFVPEKLR 309

## RESULT 7

G82293

LybB protein VC0685 [imported] - Vibrio cholerae (strain N16961 serogroup O1)  
 C/Species: Vibrio cholerae  
 C/Date: 18-Aug-2000 #sequence\_revision 20-Aug-2000 #text\_change 02-Feb-2001  
 C/Accession: G82293  
 R/Reichelberg, J.P.; Eisen, J.A.; Nelson, W.C.; Clayton, R.A.; Gwinn, M.L.; Dodson, R.J.;  
 Chardon, D.; Ermolaeva, K.D.; Vamathavan, U.; Bass, S.; Qin, H.; Dregol, I.; Sellers, P.  
 Nature 406, 477-483, 2000  
 A/Title: DNA Sequence of both chromosomes of the cholera pathogen Vibrio cholerae.  
 A/Reference number: AB2035; MUID:20406683; PMID:10952301  
 A/Accession: G82293  
 A/Status: preliminary  
 A/Molecule type: DNA  
 A/Residues: 1-320 <HEI>  
 A/Cross-references: GB:AE004154; GB:AE003852; NID:G9655115; PTDN:AAF93850.1; GSPDB:GN001;  
 A/Experimental source: serogroup O1; strain N16961; biotype El Tor  
 C/Genetics:  
 A/Gene: VC0685  
 A/Map position: 1  
 C/Superfamily: penicillin tolerance protein

## Query Match

60.8%; Score 987; DB 2; Length 320;

Best Local Similarity 61.9%; Pred. No. 7e-64;  
Matches 195; Conservative 56; Mismatches 58; Indels 6; Gaps 3;

QY 1 MOVLANPRGFCAGVDRAIEIVDAIEAFGAPLYVEHEVHNRTVVDGLKQKGAVFIEBL 60  
DB 5 MKILLANPRGFCAGVDRAISIVERALEIYOPPIYVEHEVHNRTVVDGLKQKGAVFIEBL 64  
QY 61 SDVPVGSYILFSAHGVSKVEVQAEEROLTVDPATCPVTKVMQVAKAKQGREVILIG 120  
DB 65 NEVPDDNIVIFSAHGVSKVQAEEROLTVDPATCPVTKVMQVAKAKQGREVILIG 124  
QY 121 HACHPEVEGTMGQY--ASDTGMYLVEKPEGVSLQAKVDPENLHAYGQTTLSVDETND 178  
DB 125 HACHPEVEGTMGQY--ASDTGMYLVEKPEGVSLQAKVDPENLHAYGQTTLSVDETND 182  
QY 179 MVDALEQPSIKQKDDICVATQNRQDAVHLAKISDLILVWGSPPNSNRLREIAV 238  
DB 183 VIELRRVFEKIQPKRDDICVATQNRQDAVHLAKISDLILVWGSPPNSNRLREIAV 242  
QY 239 QLGKPAVILDTYQDLKQDMLEGIEVVGTVAGASAPFVLQVQEVIDQLKAWGGETTSVENS 298  
DB 243 KLGTPGLIDCPQDIDPEWVVDQALGVTAAGASAPFVLQVQEVIDQLKAWGGETTSVENS 300  
QY 299 GIEKVVFSIPKELK 313  
DB 301 GREENMFEVPEKELQ 315

## RESULT 8

lybB protein homolog H11007 - Haemophilus influenzae (strain Rd KX20)

C/Species: Haemophilus influenzae  
C/Date: 10-Sep-1999 #sequence\_revision 10-Sep-1999 #text\_change 10-Sep-1999  
C/Accession: A64164  
R/Leischmann, R.D.; Adams, M.D.; White, O.; Claydon, R.A.; Kitchess, E.F.; Kerlayage, J.; Gocayne, J.D.; Scott, J.; Shirley, R.; Liu, L.I.; Gloddek, A.; Kellsey, J.M.; Weidman, D.M.; Brandon, R.C.; Fine, L.D.; Fitchman, J.L.; Fuhman, J.L.; Geoghegan, N.S.K.; Science 269, 496-512, 1995  
A/Authors: Gnehm, C.L.; McDonald, L.A.; Small, K.V.; Fraser, C.M.; Smith, H.O.; Venter, A./Title: Whole-genome random sequencing and assembly of Haemophilus influenzae Rd.  
A/Reference number: A64000; MUID:95350630; PMID:7542800  
A/Accession: A64164  
A/Status: nucleic acid sequence not shown; translation not shown  
A/Molecule type: DNA  
A/Residues: 1-318 <TIGR>  
A/Cross-references: GB:L42023; TIGR:H11007  
C/Genetics:  
A/Start codon: GTG  
C/Superfamily: penicillin tolerance protein

Query Match 59.7%; Score 969; DB 1; Length 318;  
Best Local Similarity 58.8%; Pred. No. 1.4e-62;  
Matches 184; Conservative 61; Mismatches 64; Indels 4; Gaps 2;

QY 1 MOVLANPRGFCAGVDRAIEIVDAIEAFGAPLYVEHEVHNRTVVDGLKQKGAVFIEBL 60  
DB 5 MKILLANPRGFCAGVDRAISIVERALEIYOPPIYVEHEVHNRTVVDGLKQKGAVFIEBL 64  
QY 61 SDVPVGSYILFSAHGVSKVEVQAEEROLTVDPATCPVTKVMQVAKAKQGREVILIG 120  
DB 65 NEVPDDNIVIFSAHGVSKVQAEEROLTVDPATCPVTKVMQVAKAKQGREVILIG 124  
QY 121 HACHPEVEGTMGQY--ASDTGMYLVEKPEGVSLQAKVDPENLHAYGQTTLSVDETND 178  
DB 125 HACHPEVEGTMGQY--ASDTGMYLVEKPEGVSLQAKVDPENLHAYGQTTLSVDETND 182  
QY 181 DAIREQPSIKQKDDICVATQNRQDAVHLAKISDLILVWGSPPNSNRLREIAV 240  
DB 183 AALKEKYPALQGHKKDDICVATQNRQDAVHLAKISDLILVWGSPPNSNRLREIAV 242  
QY 241 GKRAVILDTYQDLKQDMLEGIEVVGTVAGASAPFVLQVQEVIDQLKAWGGETTSVENS 300  
DB 243 GVSQILDEPADIQDMWFDVKTGITAGASAPFVLQVQEVIDQLKAWGGETTSVENS 300

QY 301 EKVYFSIPKELK 313  
DB 301 BGNMFEVPEKELR 313

## RESULT 9

lybB protein NME1831 [imported] - Neisseria meningitidis (strain MC58 serogroup B, strain C/Species: Neisseria meningitidis  
C/Date: 31-Mar-2000 #sequence\_revision 31-Mar-2000 #text\_change 02-Feb-2001  
C/Accession: D81038; D81982  
R/Retzlaff, H.; Saunders, N.J.; Heidelberg, J.; Jeffries, A.C.; Nelson, K.E.; Sisen, J.A.; Hickey, E.K.; Haft, D.H.; Salzberg, S.L.; White, O.; Fleischmann, R.D.; Dougherty, B.A.; ri, H.; Qin, H.; Vamathevan, J.; Gill, J.; Scarlato, V.; Masignani, V.; Pizzi, M.; Science 287, 1809-1815, 2000  
A/Authors: Grandi, G.; Sun, L.; Smith, H.O.; Fraser, C.M.; Moxon, E.R.; Rappuoli, R.; Ve A./Title: Complete genome sequence of Neisseria meningitidis serogroup B strain MC58.  
A/Reference number: A81000; MUID:20175755; PMID:10710307  
A/Accession: D81038  
A/Status: preliminary  
A/Molecule type: DNA  
A/Residues: 1-322 <T37>  
A/Cross-references: GB:A8002532; GB:A8002098; NID:g72227078; PIDN:AAF42166.1; PID:g7222708  
A/Experimental source: serogroup B, strain MC58  
R/Parhill, J.; Achtman, M.; James, K.D.; Bentley, S.D.; Churcher, C.; Klee, S.R.; Morel Nature 404, 502-506, 2000  
A/Title: Complete DNA sequence of a serogroup A strain of Neisseria meningitidis 22491.  
A/Reference number: A81775; MUID:20222556; PMID:10761919  
A/Accession: D81982  
A/Status: preliminary  
A/Molecule type: DNA  
A/Residues: 1-322 <PAR>  
A/Cross-references: GB:AL162753; GB:AL157959; NID:g7379120; PIDN:CA863914.1; PID:g737935  
A/Experimental source: serogroup A, strain 22491  
C/Genetics:  
A/Genes: lybB; NME1831; NMA0624  
C/Superfamily: penicillin tolerance protein

Query Match 59.6%; Score 968; DB 2; Length 322;  
Best Local Similarity 61.5%; Pred. No. 1.7e-62;  
Matches 192; Conservative 48; Mismatches 66; Indels 6; Gaps 2;

QY 3 IVLANPRGFCAGVDRAIEIVDAIEAFGAPLYVEHEVHNRTVVDGLKQKGAVFIEBLSD 62  
DB 6 IILANPRGFCAGVDRAISIVERALEIYOPPIYVEHEVHNRTVVDGLKQKGAVFIEBLSD 65  
QY 63 VPVGSYILFSAHGVSKVEVQAEEROLTVDPATCPVTKVMQVAKAKQGREVILIGHA 122  
DB 66 VPVGSYILFSAHGVSKVQAEEROLTVDPATCPVTKVMQVAKAKQGREVILIGHA 125  
QY 123 GHPEVEGTMGQY--GKTLVEIVGDVAKLEVRNDKLAAYGQTTLSVDETNDIAA 181  
DB 126 GHPEVEGTMGQY--GKTLVEIVGDVAKLEVRNDKLAAYGQTTLSVDETNDIAA 181  
QY 183 LREQPSIKQKDDICVATQNRQDAVHLAKISDLILVWGSPPNSNRLREIAV 242  
DB 182 LREQPSIKQKDDICVATQNRQDAVHLAKISDLILVWGSPPNSNRLREIAV 241  
QY 243 PAVILDTYQDLKQDMLEGIEVVGTVAGASAPFVLQVQEVIDQLKAWGGETTSVENS 302  
DB 242 DAYVWDNAGVILQRAAMPBKKNNVGTAGASAPFVLQVQEVIDQLKAWGGETTSVENS 299  
QY 303 KVPVFSIPKELK 314  
DB 300 STVPLPEKELR 311

## RESULT 10

drug tolerance protein X52416 [imported] - Xylella fastidiosa (strain 945c)  
C/Species: Xylella fastidiosa



QY 181 DALREDFPSIKKQKDDICATONRODAVHDLAKISDLILVWSSPNSSNSNRLREIAVQL 240  
 DB 187 ALRFRPFGIAAPKXEDICATNRQDAVHDLAKISDLILVWSSPNSSNSNRLREIAVQL 246  
 QY 241 G-KPAILIDYQDLKQWLEGEIVYGTAGASAPVILVQEVYDQKMGCEITTSVRENSG 239  
 DB 247 GARDALIDPASGIDMTWFGISRVGLTAGASAPEDLVQVDAISA--RFDITVELVE 304  
 QY 300 IEEKVFSIPKEL 312  
 DB 305 ARRTITKRLRL 317

## RESULT 13

AH2671

penicillin tolerance protein (imported) - Agrobacterium tumefaciens (strain C58, Dupont)

C:Species: Agrobacterium tumefaciens  
 C>Date: 11-Jan-2002 #sequence\_revision 11-Jan-2002 #text\_change 18-Nov-2002

C:Accession: AH2671

R:Wood, D.W.; Setubal, J.C.; Kaul, R.; Monks, D.; Chen, L.; Wood, G.B.; Chen, Y.; Woo, L.; Karp, P.; Romero, P.; Grant, C.; Guenther, D.; Kuttyavin, T.; Levy, R.; Li, M.; McClellan, S.  
 Science 294, 2317-2323, 2001

A:Authors: Yoo, H.; Tao, Y.; Biddle, P.; Jung, M.; Krespan, W.; Perry, M.; Gordon-Kamm, S.E.M.  
 A:Title: The Genome of the Natural Genetic Engineer Agrobacterium tumefaciens C58.  
 A:Reference number: AB2577; MUID:21608550; PMID:11743393

A:Accession: AH2671

A>Status: preliminary

A:Molecule type: DNA

A:Residues: 1-348 <KUR>

A:Cross-references: GB:AB006688; PIDN:AAL41790.1; PID:gl7739144; GSPDB:GN00186

A:Experimental source: strain C58 (Dupont)

C:Genetics:

A:Gene: lycB

A:Map position: circular chromosome

C:Superfamily: penicillin tolerance protein

Query Match 46.6%; Score 756; DB 2; Length 348;  
 Best Local Similarity 50.5%; Pred. No. 3.7e-47;

Matches 161; Conservative 54; Mismatches 90; Indels 14; Gaps 6;

QY 1 MOVLAMPFGCAGVDRAIEVDQAIKAFGAPVYRHEVHNRTVVDGLKQKAVFIEEL 60  
 DB 10 LTRLCGRPGFCAGVDRAIQIVLALKAQAPVYRHEVHNRTVVDGLKQKAVFIEEL 69  
 QY 61 SDVPEV--GSYLIFSAHGSKEVQAEERQLTVPDAPCTPLTKTMQVAKAKQREVI 117  
 DB 70 HEIPAEHREQPVFSAHGVKSPVEDAQARNLFYLDATCPVSKVKQAKRQRLGRHV 129  
 QY 118 LIGHAGHPVEGTMGQYEKCTEGGIGYLVETPEDEVNLIKNNPNDLAYVOTLLSMTDK 177  
 DB 130 LIGHAGHPVIGTQGLPE---GTVSLVETVEDAGVYEPVDRENIGFVOTLLSVDDTA 185  
 QY 178 VMVDALREDFPSIKKQKDDICATONRODAVHDLAKISDLILVWSSPNSSNSNRLREIA 217  
 DB 186 GVIALRQERFPFPAIQAPADSICVATNRQDAVQQAAPGCDLFIYVGAAPNSNSNRLREIA 245  
 QY 238 VQUG-KPAILIDYQDLKQWLE--GEIVYGTAGASAPVILVQEVYDQKMGCEITTSV 294  
 DB 246 LRAAKHSVILVQASAEI--DMNEIGDIRTVGLSAGASAPVILVQEVYDQKMGCEITTSV 303  
 QY 295 RENSGIEKVFSPKEL 313  
 DB 304 AVT--VETHEFLVNRRLR 320

## RESULT 14

P97453

lyb protein (AB006203) (imported) - Agrobacterium tumefaciens (strain C58, Cereon)

C:Species: Agrobacterium tumefaciens  
 C>Date: 30-Sep-2001 #sequence\_revision 30-Sep-2001 #text\_change 18-Nov-2002  
 C:Accession: P97453

R:Goodner, B.; Hinkle, G.; Gattung, S.; Miller, N.; Blanchard, M.; Quorllo, B.; Goldman, A.; Liu, F.; Mollan, C.; Allinger, M.; Doughy, D.; Scott, C.; Lappes, C.; Markelz, B.; Science 294, 2323-2328, 2001  
 A:Title: Genome Sequence of the Plant Pathogen and Biotechnology Agent Agrobacterium tumefaciens C58  
 A:Reference number: A97359; MUID:21608551; PMID:11743194  
 A:Accession: P97453  
 A>Status: preliminary  
 A:Molecule type: DNA  
 A:Residues: 1-348 <KUR>  
 A:Cross-references: GB:AB007869; PIDN:AAK6583.1; PID:gl5155749; GSPDB:GN00186  
 C:Genetics: AGR\_C1414  
 A:Gene: AGR\_C1414  
 A:Map position: circular chromosome  
 C:Superfamily: penicillin tolerance protein

Query Match 46.6%; Score 756; DB 2; Length 348;  
 Best Local Similarity 50.5%; Pred. No. 3.7e-47;

Matches 161; Conservative 54; Mismatches 90; Indels 14; Gaps 6;

QY 1 MOVLAMPFGCAGVDRAIEVDQAIKAFGAPVYRHEVHNRTVVDGLKQKAVFIEEL 60  
 DB 10 LTRLCGRPGFCAGVDRAIQIVLALKAQAPVYRHEVHNRTVVDGLKQKAVFIEEL 69  
 QY 61 SDVPEV--GSYLIFSAHGSKEVQAEERQLTVPDAPCTPLTKTMQVAKAKQREVI 117  
 DB 70 HEIPAEHREQPVFSAHGVKSPVEDAQARNLFYLDATCPVSKVKQAKRQRLGRHV 129  
 QY 118 LIGHAGHPVEGTMGQYEKCTEGGIGYLVETPEDEVNLIKNNPNDLAYVOTLLSMTDK 177  
 DB 130 LIGHAGHPVIGTQGLPE---GTVSLVETVEDAGVYEPVDRENIGFVOTLLSVDDTA 185  
 QY 178 VMVDALREDFPSIKKQKDDICATONRODAVHDLAKISDLILVWSSPNSSNSNRLREIA 217  
 DB 186 GVIALRQERFPFPAIQAPADSICVATNRQDAVQQAAPGCDLFIYVGAAPNSNSNRLREIA 245  
 QY 238 VQUG-KPAILIDYQDLKQWLE--GEIVYGTAGASAPVILVQEVYDQKMGCEITTSV 294  
 DB 246 LRAAKHSVILVQASAEI--DMNEIGDIRTVGLSAGASAPVILVQEVYDQKMGCEITTSV 303  
 QY 295 RENSGIEKVFSPKEL 313  
 DB 304 AVT--VETHEFLVNRRLR 320

## RESULT 15

AB3434

lyb protein (imported) - Brucella melitensis (strain 16M)

C:Species: Brucella melitensis  
 C>Date: 01-Feb-2002 #sequence\_revision 01-Feb-2002 #text\_change 15-Feb-2002

C:Accession: AB3434

R:Belvecchio, V.G.; Kapral, V.; Redkar, R.U.; Patre, G.; Muier, C.; Joe, T.; Ivanova, M.; Mazur, M.; Goldsman, E.; Selkov, E.; Elzer, P.H.; Hagjue, S.; O'Callaghan, D.; Letes, Proc. Natl. Acad. Sci. U.S.A. 99, 443-448, 2002

A:Title: The genome sequence of the facultative intracellular pathogen Brucella melitensis  
 A:Reference number: AD3252; PMID:11756688

A:Accession: AB3434

A>Status: preliminary

A:Molecule type: DNA

A:Residues: 1-346 <KUR>  
 A:Cross-references: GB:AB008917; PIDN:AAL52640.1; PID:gl7983462; GSPDB:GN00190

A:Experimental source: strain 16M

C:Genetics:

A:Gene: BME11459

A:Map position: I

C:Superfamily: penicillin tolerance protein

Query Match 46.2%; Score 750; DB 2; Length 346;  
 Best Local Similarity 47.7%; Pred. No. 1e-46;

Matches 154; Conservative 59; Mismatches 88; Indels 22; Gaps 5;

QY 1 MOVLAMPFGCAGVDRAIEVDQAIKAFGAPVYRHEVHNRTVVDGLKQKAVFIEEL 60  
 DB 8 LTRLCGRPGFCAGVDRAIQIVLALKAQAPVYRHEVHNRTVVDGLKQKAVFIEEL 67

```

QY 61 SDPVV---GSYLLIFSAHVSKEVVOQSAEROLFVFDATCELVTKVHMVAKHAKOGREVI 117
Db 68 DEIPAHARNQPVVPSAHGVKSVPADAEAKNLFYLDATCPVSKVHKQAMRHQRLGRHVI 127
QY 118 LIGHAGHPVEGTMGQYEKCTBGGGILVETPEDVBNLKVNNPNDLAVYTOFTLSMTDX 177
Db 128 LIGHSGHPVEVIGTMGQ---LPDGAVTLLIETVEDAHTCHFDDEDNLGFTVOTTLISVDDTA 183
QY 178 VMVDALREOFPSIKEQKDDICVATORODAVHDLAKISDILLVGSPPNSNSNRLEITA 237
Db 184 GIITKELOAKRFPNLAAPAAESICVATTNRQDAVRAAAGCDLFLVGA PNSSNSKRLVEVA 243
QY 238 VQLG-KPAYLIDTYQDLKQDMLBEGIEVVGATAGASAPVWQEVTDOLKANG-----GE 290
Db 244 EKAGARMSMLVQRAEDIEWQIGDISVVGLSAGASAPEIIVDEIIDAFAKARPDVXIELAE 303
QY 291 TTSVRENSGIEEKVPSIKEK 313
Db 304 TTVETEN-----FLVNRREIR 318

```

Search completed: January 29, 2004, 15:55:34  
 Job time : 10.3566 secs

GenCore version 5.1.6  
Copyright (c) 1993 - 2004 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: January 29, 2004, 15:41:08 ; Search time 5.98823 seconds  
(without alignments)  
2497.314 Million cell updates/sec

Title: US-09-941-947A-18  
Perfect score: 1623  
Sequence: 1 MOIYLANPRGFCAGVDRAIE.....GIEKVFESIPRELKKNQQA 318

Scoring table: BLOSUM62  
Gapop 10.0 , Gapext 0.5

Searched: 127863 seqs, 47026705 residues

Total number of hits satisfying chosen parameters: 127863

Minimum DB seq length: 0  
Maximum DB seq length: 200000000

Post-Processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Database : SwissProt\_41.\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

## SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	1064.5	65.6	316	1 ISPH_ACTICA	Q9xj0 acinetobact
2	1028.5	63.4	314	1 ISPH_PSEAE	Q9hmt7 pseudomonas
3	1019.5	62.8	317	1 ISPH_YRPE	P58680 yersinia pe
4	1001	61.7	316	1 ISPH_SALT1	P58678 salicicellul
5	996	61.4	316	1 ISPH_ECOLI	P22655 escherichia
6	996	61.4	316	1 ISPH_SALTY	P58679 salicicellul
7	989.5	61.0	321	1 ISPH_VIBPA	Q87877 vibrio para
8	989	60.9	320	1 ISPH_VIBBU	Q87878 vibrio para
9	987	60.8	316	1 ISPH_VIBCH	Q9xk44 vibrio chol
10	980	60.4	316	1 ISPH_XANAC	Q8p17 xanthomonas
11	970	59.8	316	1 ISPH_XANCP	Q8p094 xanthomonas
12	969	59.7	314	1 ISPH_HAEIN	P44976 haemophilus
13	969	59.7	314	1 ISPH_PASMT	P57960 pasteurella
14	968	59.6	322	1 ISPH_NEIMA	Q91c39 neisseria m
15	967	59.6	316	1 ISPH_XYLFA	Q9p89 xyloella fas
16	949.5	58.5	324	1 ISPH_RALSO	P58677 ralsstonia s
17	866	53.4	315	1 ISPH_WIGBR	Q8d122 wigglewort
18	858	52.9	319	1 ISPH_BUCAR	P57247 buchnera ap
19	826.5	50.9	319	1 ISPH_CAUCR	Q9a345 caulobacter
20	824.5	50.8	313	1 ISPH_BUCAP	Q8x924 buchnera ap
21	813.5	50.1	313	1 ISPH_BURPS	Q9f10 burkholderi
22	791	48.7	338	1 ISPH_STROO	Q9f10 streptomyce
23	756	46.6	348	1 ISPH_AGRYS	P58673 agrobacteri
24	755	46.5	325	1 ISPH_CORGL	Q9hmt2 corynobacte
25	752.5	46.4	350	1 ISPH_RHIME	Q9xj02 rhizobium m
26	750	46.2	346	1 ISPH_BRUME	Q9xj02 rhizobium m
27	743.5	45.8	329	1 ISPH_MYCTU	O54409 mycobacteri
28	740	45.6	336	1 ISPH_RHILU	O54409 mycobacteri
29	733	45.2	335	1 ISPH_MYCTU	O54458 mycobacteri
30	717	44.2	332	1 ISPH_MYCTE	Q9x781 mycobacteri
31	667	41.1	307	1 ISPH_CHLTR	Q8a867 chlamydia t
32	663	40.9	310	1 ISPH_CHEPN	Q9x6p2 chlamydia p
33	661	40.7	308	1 ISPH_CHEMU	Q9p159 chlamydia m

## ALIGNMENTS

RESULT 1	ID	ISPH_ACTICA	STANDARD	PRT	316 AA.
AC	Q9xj0:	28-FEB-2003 (rel. 41, last sequence update)			
DT	28-FEB-2003 (rel. 41, last sequence update)				
DE	ISPH protein.				
GN	ISPH OR LYTB.				
OS	Acinetobacter calcoaceticus.				
OC	Bacteria; Proteobacteria; Gammaproteobacteria; Pseudomonadales;				
OX	Moraxellaceae; Acinetobacter.				
RM	NCBI_TaxID=471;				
RP	SEQUENCE FROM N.A.				
RC	STRAIN=BD413 / ADP1;				
RX	MEDLINE=20224825; PubMed=10763755;				
RA	Herzberg C., Friedrich A., Averhoff B.;				
RT	"Comb, a novel competence gene required for natural transformation of				
RL	Acinetobacter sp. BD413: identification, characterization and analysis				
RT	of growth phase-dependent regulation.";				
Arch.	Microbiol. 173:220-228(2000).				
CC	-1- FUNCTION: Converts 1-hydroxy-2-methyl-2-(E)-butenyl 4-diphosphate				
CC	into isopentenyl diphosphate (IPP) and dimethylallyl diphosphate				
CC	(DMAPP) (by similarity).				
CC	-1- PATHWAY: Nonmevalonate terpenoid biosynthesis pathway; seventh				
CC	(last) step.				
CC	-1- SIMILARITY: BELONGS TO THE ISPH FAMILY.				
CC	This SWISS-PROT entry is copyright. It is produced through a collaboration				
CC	between the Swiss Institute of Bioinformatics and the EMBL outstation.				
CC	the European Bioinformatics Institute. There are no restrictions on its				
CC	use by non-profit institutions as long as its content is in no way				
CC	modified and this statement is not removed. Usage by and for commercial				
CC	entities requires a license agreement (See <a href="http://www.isb-sib.ch/announce/">http://www.isb-sib.ch/announce/</a>				
CC	or send an email to <a href="mailto:license@isb-sib.ch">license@isb-sib.ch</a> ).				
DR	EMBL: AF077189; AAC55804.1;				
DR	HANAB: MF_001911; -- 1.				
DR	InterPro: IPR003451; LYTB.				
DR	Pfam: PF02401; LYTB; 1.				
DR	TIGRFAMs: TIGR00216; isph_lytb; 1.				
KM	ISOPrene biosynthesis.				
SC	SEQUENCE 316 AA; 34943 MW; 29F071CER2B9A762 CRC64;				
Qy	Query Match	65.6%;	Score 1064.5;	DB 1;	Length 316;
Qy	Best Local Similarity	65.2%;	Pred. No. 2.1e-66;		
Db	Matches 204;	Conservative 48;	Mismatches 58;	Indels 3;	Gaps 2;
Qy	1 MOIYLANPRGFCAGVDRAIEIVDAIPAPGVYRHHVNNRVVNGLKQKQGVFIIEEL 60				
Db	1 MEIYLANPRGFCAGVDRAIAIVNAALCEFNPIYVRAEIVHAKFVDDDLQKQGAIFVDEL 60				
Qy	61 SDPEVGSYLLFSAAGVSKVQQAEEBRLTFPATCELVTKVHQAQAKAQKQREVLILG 120				
Db	61 DQVEEDSDIVIFSAAGVSKVQQAEEBRLTFPATCELVTKVHQAQAKAQKQREVLILG 120				

QY 121 HAGHEVEGTMGQYKCTEGGGIYLVETPEDEVANLKNPNPDLYAVYOTTLSTMTDKRMV 180  
 DB 121 HSHHPEVEGTMGQYK-SKGGHLYVEDEADVALLVNEPELAFYOTTLSTMTDKRMV 179  
 QY 181 DALREOPFSIKQKDDICATQNRQDAVADLAKISDLILVGSPPNSNSRLREIAVOL 240  
 DB 180 DALRTEFPQIQGPRKNDICATQNRQDAVADLAKISDLILVGSPPNSNSRLREIAVOL 239  
 QY 241 GKPAYLIDTYODLQMDLEGIEVGVGTAGASAPVLYVQEVYDOLKAMGETTSYRENSGI 300  
 DB 240 GKPAYLIDTYODLQMDLEGIEVGVGTAGASAPVLYVQEVYDOLKAMGETTSYRENSGI 297  
 QY 301 BEKVFSIPKELK 313  
 DB 298 EENITFSIPKELR 310

## RESULT 2

ISPH\_PSEAE STANDARD; PRT; 314 AA.  
 AC 09HVM7;  
 DT 16-OCT-2001 (Rel. 40, Created)  
 DT 16-OCT-2001 (Rel. 40, Last sequence update)  
 DT 28-FEB-2003 (Rel. 41, Last annotation update)  
 DE ISPH protein.  
 GN ISPH OR LYTB OR PA4557.  
 OS Pseudomonas aeruginosa.  
 OC Bacteria; Proteobacteria; Gammaproteobacteria; Pseudomonadales;  
 OC Pseudomonadaceae; Pseudomonas.  
 OX NCBI\_TaxID=287;  
 RN [1]

RP SEQUENCE FROM N.A.  
 RC STRAIN=ATCC 15692 / PA01;  
 RA MEDLINE=20437337; PubMed=10984043;  
 RA Stover C.K., Pham X.-Q.T., Erwin A.L., Mizoguchi S.D., Warren P.,  
 RA Hickey M.J., Brinkman F.S.L., Huffnagle W.O., Kowalik D.J., Lagrou M.,  
 RA Garber R.L., Goltry L., Tolentino E., Westbrook-Wadman S., Yan Y.,  
 RA Brody L.L., Collier S.N., Folger K.R., Kas A., Laidig K., Lim R.K.,  
 RA Smith K.A., Spencer D.H., Wong G.K.-S., Wu Z., Paulsen I.T.,  
 RA Reiter U., Sater M.H., Hancock R.E.W., Lory S., Olson M.V.,  
 RA "Complete genome sequence of Pseudomonas aeruginosa PA01, an  
 RT opportunistic pathogen.";  
 RT Nature 406:959-964 (2000).  
 RL Nature 406:959-964 (2000).

CC - FUNCTION: Converts 1-hydroxy-2-methyl-2-(E)-butenyl 4-diphosphate  
 CC into isopentenyl diphosphate (IPP) and dimethylallyl diphosphate  
 CC (DMAPP) (by similarity).  
 CC - PATHWAY: Nonmevalonate terpenoid biosynthesis pathway; seventh  
 CC (last) step.

CC - SIMILARITY: BELONGS TO THE ISPH FAMILY.

CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (see <http://www.isb-sib.ch/announce/>  
 CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).

CC EMBL: AE004863; AAC07945.1; --  
 CC PIR: G83076; G83076.  
 CC HAMAP: MF\_00191; -- 1.  
 CC InterPro: IPR003451; IYB.  
 CC Pfam: PF02401; LYTB.1.  
 CC TIGRFAMs: TIGR00216; isph\_lytb.1.  
 CC isoprene biosynthesis; Complete proteome.  
 CC KW SEQUENCE 314 AA; 34762 MW; 483E07B68663EEB6 CRC64;

Query Match 63.4%; Score 1028.5; DB 1; Length 314;  
 Best Local Similarity 63.3%; Pred. No. 7.5e-66;  
 Matches 198; Conservative 49; Mismatches 63; Indels 3; Gaps 2;

QY 1 MOYILANRGACAGDRAILEIVDAIEARGAIRYRHEVYHNRITYVGLKQKGVFIIEEL 60

DB 1 MOYILANRGACAGDRAILEIVDAIEARGAIRYRHEVYHNRITYVGLKQKGVFIIEEL 60  
 QY 61 SDVPGSLTISAGSSEKVEQGEARROLYTFDQTCPLYTVHMQVAKHAGQREVIILIG 120  
 DB 61 DQVPMNVAVIIPASAGVSQAVKKEARGLKVFDTCPLYTVHMEVAVSRDGEHCVALIG 120  
 QY 121 HAGHEVEGTMGQYKCTEGGGIYLVETPEDEVANLKNPNPDLYAVYOTTLSTMTDKRMV 180  
 DB 121 HSHHPEVEGTMGQYD-ASNGALYLVDEADVALLVNEPELAFYOTTLSTMTDKRMV 179  
 QY 181 DALREOPFSIKQKDDICATQNRQDAVADLAKISDLILVGSPPNSNSRLREIAVOL 240  
 DB 180 DALRAKFPQIQGPRKNDICATQNRQDAVADLAKISDLILVGSPPNSNSRLREIAVOL 239  
 QY 241 GKPAYLIDTYODLQMDLEGIEVGVGTAGASAPVLYVQEVYDOLKAMGETTSYRENSGI 300  
 DB 240 GKPAYLIDTYODLQMDLEGIEVGVGTAGASAPVLYVQEVYDOLKAMGETTSYRENSGI 297  
 QY 301 BEKVFSIPKELK 313  
 DB 298 EENITFSIPKELR 310

## RESULT 3

ISPH\_YERPE STANDARD; PRT; 317 AA.  
 AC P56880;  
 DT 28-FEB-2003 (Rel. 41, Created)  
 DT 28-FEB-2003 (Rel. 41, Last sequence update)  
 DT 28-FEB-2003 (Rel. 41, Last annotation update)  
 DE ISPH protein.  
 GN ISPH OR LYTB OR YPO0477 OR Y3697.  
 OS Yersinia pestis.  
 OC Bacteria; Proteobacteria; Gammaproteobacteria; Enterobacteriales;  
 OC Enterobacteriaceae; Yersinia.  
 OX NCBI\_TaxID=632;  
 RN [1]

RP SEQUENCE FROM N.A.  
 RC STRAIN=KIM5 / Biovar Mediaevalis;  
 RA MEDLINE=21470413; PubMed=11586360;  
 RA Partridge J., Wren B.W., Thomson N.R., Tibball R.W., Holden M.T.G.,  
 RA Partridge J.B., Sebalhia M., James K.D., Churcher C., Mungall K.L.,  
 RA Baker S., Basham D., Bentley S.D., Brooks K., Cerdano-Terriza A.M.,  
 RA Chillingworth T., Cronin A., Davies R.W., Davis P., Dougan G.,  
 RA Feltham S., Hamlin N., Holroyd S., Jagers K., Kariyeh A.V.,  
 RA Leather S., Moule S., Oyston P.C.F., Quail M., Rutherford K.,  
 RA Simmonds M., Skelton J., Stevens K., Whitehead S., Barrett B.G.,  
 RA "Genome sequence of Yersinia pestis, the causative agent of plague.";  
 RT Nature 413:523-527 (2001).  
 RL Nature 413:523-527 (2001).

CC - FUNCTION: Converts 1-hydroxy-2-methyl-2-(E)-butenyl 4-diphosphate  
 CC into isopentenyl diphosphate (IPP) and dimethylallyl diphosphate  
 CC (DMAPP) (by similarity).  
 CC - PATHWAY: Nonmevalonate terpenoid biosynthesis pathway; seventh  
 CC (last) step.

CC - SIMILARITY: BELONGS TO THE ISPH FAMILY.

CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial

RC STRAIN=Ty2 / ATCC 700931.

Enterobacteriaceae; Escherichia.

OX NCBI\_TaxID=562, 217992, 83334;  
 RN [1] SEQUENCE FROM N.A., AND CHARACTERIZATION.  
 RX MEDLINE=21819424; PubMed=11818558;  
 RA Reichelt F., Hecht S., Gaertner K., Adam P., Krieger C., Anslinger S.,  
 RA Arigoni D., Bachert A., Easenteich W.,  
 RT "Studies on the nonmevalonate terpene biosynthetic pathway: metabolic  
 RT role of Isph (lytB) protein.";  
 RL Proc. Natl. Acad. Sci. U.S.A. 99:1158-1163(2002).  
 RN [2]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=K12;  
 RX MEDLINE=91187617; PubMed=2011499;  
 RA Bouvier J., Stragler P.,  
 RT "Nucleotide sequence of the isp-dapB interval in Escherichia coli.";  
 RL Nucleic Acids Res. 19:180-180(1991).  
 RN [3]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=K12;  
 RX MEDLINE=92334977; PubMed=1630901;  
 RA Yura T., Mori H., Nagai H., Nagata T., Ishihama A., Fujita N.,  
 RA Isono K., Mizobuchi K., Nakata A.,  
 RT "Systematic sequencing of the Escherichia coli genome: analysis of  
 RT the 0-2.4 min region.";  
 RL Nucleic Acids Res. 20:3305-3308(1992).  
 RN [4]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=K12 / MG1655;  
 RX MEDLINE=97426617; PubMed=9278503;  
 RA Blattner F.R., Plunkett G. III, Bloch C.A., Perna N.T., Burland V.,  
 RA Riley M., Collado-Vides J., Glasner J.D., Rode C.K., Mayhew G.F.,  
 RA Gregor J., Davis N.W., Kirkpatrick H.A., Goeden M.A., Rose D.J.,  
 RA Mau B., Shao Y.,  
 RT "The complete genome sequence of Escherichia coli K-12.";  
 RL Science 277:1234-1244(1997).  
 RN [5]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=O6:H1 / CFT073 / ATCC 700928;  
 RX MEDLINE=22388234; PubMed=12471157;  
 RA Welch R.A., Burland V., Plunkett G. III, Redford P., Roesch P.,  
 RA Rasko D., Buckles E.L., Liou S.-R., Boutin A., Hackett J., Stroud D.,  
 RA Mayhew G.F., Rose D.J., Zhou S., Schwartz D.C., Perna N.T.,  
 RA Mobley H.L.T., Donnenberg M.S., Blattner F.R.,  
 RT "Extensive mosaic structure revealed by the complete genome sequence  
 RT of uropathogenic Escherichia coli.";  
 RL Proc. Natl. Acad. Sci. U.S.A. 99:17020-17024(2002).  
 RN [6]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=O157:H7 / EDL933 / ATCC 700927;  
 RX MEDLINE=21074935; PubMed=11206551;  
 RA Perna N.T., Plunkett G. III, Burland V., Mau B., Glasner J.D.,  
 RA Rose D.J., Mayhew G.F., Evans P.S., Gregor J., Kirkpatrick H.A.,  
 RA Postel G., Hackett J., Klink S., Boutin A., Shao Y., Miller L.,  
 RA Grobbeck E.J., Davis N.W., Lam A., Dimalanta E.T., Potamous K.,  
 RA Apodaca J., Anantharaman T.S., Lin J., Yen G., Schwartz D.C.,  
 RA Welch R.A., Blattner F.R.,  
 RT "Genome sequence of enterohaemorrhagic Escherichia coli O157:H7.";  
 RL Nature 409:529-533(2001).  
 RN [7]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=O157:H7 / RIMD 0509952;  
 RX MEDLINE=21156231; PubMed=11258796;  
 RA Hayashi T., Makino K., Onishi M., Kurokawa K., Ishii K., Yokoyama K.,  
 RA Han C.-G., Ohtsubo E., Nakayama K., Murata T., Tanaka M., Tobe T.,  
 RA Iida T., Takami H., Honda T., Sasekawa C., Ogasawara N., Yasunaga T.,  
 RA Kubota S., Shida T., Hattori M., Shinagawa H.,  
 RT "Complete genome sequence of enterohemorrhagic Escherichia coli  
 RT O157:H7 and genomic comparison with a laboratory strain K-12.";  
 RL DNA Res. 8:11-22(2001).  
 RN [8]  
 RP IDENTIFICATION.  
 RX MEDLINE=93163053; PubMed=8432714;  
 RA Gustafson C.E., Kaul S., Ishiguro E.E.,

RT "Identification of the Escherichia coli lytB gene, which is involved  
 RT in penicillin tolerance and control of the stringent response.";  
 RL J. Bacteriol. 175:1203-1205(1993).  
 RN [9]  
 RP PATHWAY.  
 RC STRAIN=K12 / ATCC 23716 / DSM 498;  
 RX MEDLINE=2111595; PubMed=11418107;  
 RA Altincicek B., Kollas A.-K., Eberl M., Wiesner J., Sanderbrand S.,  
 RA Hintz M., Beck E., Jomaa H.,  
 RT "lytB, a novel gene of the 2-C-methyl-D-erythritol 4-phosphate pathway  
 RT of isoprenoid biosynthesis in Escherichia coli.";  
 RL FEBS Lett. 499:37-40(2001).  
 RN [10]  
 RP PATHWAY.  
 RX MEDLINE=21574179; PubMed=11717301;  
 RA McAteer S., Coulson A., McLennan N., Masters M.,  
 RT "The lytB gene of Escherichia coli is essential and specifies a  
 RT product needed for isoprenoid biosynthesis.";  
 RL J. Bacteriol. 183:7403-7407(2001).  
 CC - FUNCTION: Converts 1-hydroxy-2-methyl-2-(3)-butenyl 4-diphosphate  
 CC into isopentenyl diphosphate (IPP) and dimethylallyl diphosphate  
 CC (DMAPP). Is also involved in penicillin tolerance and control of  
 CC the stringent response. Seems to directly or indirectly interact  
 CC with rda to maintain it in an inactive form during normal growth.  
 CC - PATHWAY: Nonmevalonate terpenoid biosynthesis pathway; seventh  
 CC (last) step.  
 CC - SIMILARITY: BELONGS TO THE ISPH FAMILY.  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (see <http://www.isb-sib.ch/announce/>  
 CC or send an email to [license@sib-sib.ch](mailto:license@sib-sib.ch)).  
 CC -----  
 DR EMBL, AY062212; AAL38655.1; -  
 DR EMBL, X54945; CAA38707.1; -  
 DR EMBL, D10483; BAB96598.1; -  
 DR EMBL, AE000113; AAC73140.1; -  
 DR EMBL, AE016755; AAN78533.1; ALT\_INIT.  
 DR EMBL, AB005179; AAG54331.1; -  
 DR EMBL, AP002550; BAB33455.1; -  
 DR PIR, G85483; G85483.  
 DR PIR, H90632; H90632.  
 DR PIR, J60403; J60403.  
 DR Ecogen; BGI1081; isph.  
 DR HAMAP; MF 00191; -; 1.  
 DR InterPro; IPR003451; lytB.  
 DR Pfam; PF02401; lytB; 1  
 DR TIGRfam; TIGR00216; isph\_lytB; 1  
 KM Isoprene biosynthesis; Complete proteome.  
 SQ SEQUENCE 316 AA; 34774 MW; 0E7B378BD49AD771 CRC64;  
 Query Match 61.4%; Score 996; DB 1; Length 316;  
 Best local similarity 61.3%; Pred. NO. 1.5e-63;  
 Matches 192; Conservativity 59; Mismatches 56; Indels 4; Gaps 2;  
 QY 1 MOIYANPRGFCAGYDRAIEIYDQAIKFAPIYRREYVNRRTYDGLKQKGAIFISRL 60  
 DB 1 MOILANPRGFCAGYDRAISIVENALAYGAPITVRHVNRVYVDSIREGALFISQI 60  
 QY 61 SVVPAGSYLIFSAHVSKEVQOAEERQLTVPDACPVTQKVMQVAKAKQGRVILIG 120  
 DB 61 SEVPAGALILIFSAHVSQAVRKAASRLVFDACPVLTKVMHVAASRGEISILIG 120  
 QY 121 FAGHPEVEGTNGQYKCTBGGGSIYVETPEDEVNRINKVNPNDLAYVYOTTSMTDKVMV 180  
 DB 121 FAGHPEVEGTNGQYKCTBGGGSIYVETPEDEVNRINKVNPNDLAYVYOTTSMTDKVMV 180  
 QY 122 FAGHPEVEGTNGQYKCTBGGGSIYVETPEDEVNRINKVNPNDLAYVYOTTSMTDKVMV 180  
 DB 122 FAGHPEVEGTNGQYKCTBGGGSIYVETPEDEVNRINKVNPNDLAYVYOTTSMTDKVMV 180  
 QY 181 DALRQPFPSIKQKKDDICVATQNRQDAVHDAKISDILIVGSPGNSNRRLREIAVOE 240  
 DB 179 DALRQPFPSIKQKKDDICVATQNRQDAVHDAKISDILIVGSPGNSNRRLREIAVOE 240

QY 241 GKPAYLIDTYODLKQDMLGIEVWGTAGASAPVLYQVBIQDOKAMGSETTSVRENSGI 300  
 DB 239 GKBAFLIDDAKDQGEVWVKVCKCVCVAGASAPDILVQVYVAKQUGGSEALPLE--GR 296  
 QY 301 EKVYVPSIPKELK 313  
 DB 297 ENIVVEVPEKELR 309

## RESULT 6

ISPH\_SALTY STANDARD; PRT; 316 AA.  
 ID ISPH\_SALTY  
 AC PS8679;  
 DT 28-FEB-2003 (Rel. 41, Created)  
 DT 28-FEB-2003 (Rel. 41, Last sequence update)  
 DT 28-FEB-2003 (Rel. 41, Last annotation update)  
 DE IspH Protein  
 GN IspH OR LYTB OR STM0049.  
 OS Salmomella typhimurium.  
 CC Bacteria; Proteobacteria; Gammaproteobacteria; Enterobacteriales;  
 CC Enterobacteriaceae; Salmomella.  
 CX NCBI\_TaxID=602;  
 RX NCBI\_TaxID=602;

RP SEQUENCE FROM N.A.  
 RC STRAIN=L72 / SCS0412 / ATCC 700720;  
 RX MEDLINE=21534948; PubMed=11677609;  
 RA McDaniel M., Sanderson K.E., Spiech J., Clifton S.W., Latreille P.,  
 RA Courtney L., Porwollik S., Ali J., Dante M., Du P., Hou S., Layman D.,  
 RA Leonard S., Nguyen C., Scott K., Holmes A., Grevell N., Mulvaney E.,  
 RA Ryan B., Sun H., Flores L., Miller W., Stoneking T., Neam M.,  
 RA Waterston R., Wilson R.K.;  
 RT "Complete genome sequence of Salmomella enterica serovar Typhimurium  
 L72.";  
 RL Nature 413:852-856(2001).  
 CC -1- FUNCTION: Converts 1-hydroxy-2-methyl-2-(E)-butenyl 4-diphosphate  
 into isopentenyl diphosphate (IPP) and dimethylallyl diphosphate  
 (DMAPP) (By similarity).  
 CC -1- PATHWAY: Nonmevalonate terpenoid biosynthesis pathway; seventh  
 CC (last) step.  
 CC -1- SIMILARITY: BELONGS TO THE ISPH FAMILY.  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
 CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
 CC EMBL; AF008695; AL13013.1;  
 DR Stgyene; SG7777; ISPH.  
 DR HAMAP; MF\_00191; -; 1.  
 DR InterPro; IPR003451; LYCB.  
 DR Pfam; PF02401; LYTB; 1.  
 DR TIGRFAMs; TIGR00216; isph LYTB; 1.  
 DR Isoprene biosynthesis; Complete proteome.  
 KW SEQUENCE 316 AA; 34527 MW; 953151A411FB522 CRC64;

Query Match 61.4%; Score 996; DB 1; Length 316;  
 Best Local Similarity 62.0%; Pred. No. 1.5e-63;  
 Matches 194; Conservative 54; Mismatches 61; Indels 4; Gaps 2;

QY 1 MOYLANPFGFCAGVDALIEYDQALFAFAPITYRHHVHNRTVYDGLKQKGAFFIEEL 60  
 DB 1 MOYLANPFGFCAGVDALIEYDQALFAFAPITYRHHVHNRTVYDGLKQKGAFFIEEL 60  
 QY 61 SDVAVGSYLLIFSAHGVSKVQAEERQITVPDATCPLVTKVHMQVAKHAKGSEVILIG 120  
 DB 61 SEVPDGLILFSAHGVSKVQAEERQITVPDATCPLVTKVHMQVAKHAKGSEVILIG 120  
 QY 121 HAGHPEVGTMGQYKCTBGSGIYIVETPEDVRLKVNPNDAVYOTTLISMTDTKYV 180  
 DB 121 HAGHPEVGTMGQYKCTBGSGIYIVETPEDVRLKVNPNDAVYOTTLISMTDTKYV 180

DB 121 HAGHPEVGTMGQYS--NPEGWYLVESPEDVRLVKNKGLSPMTQTLISVDTSDVI 178  
 QY 181 DALREPFSEIKQKCDICATQNRDAYHDAKISDLILVSGSPNSNSNRLEIAVOL 240  
 DB 179 DALRKRFPIVGPBKDDICATTNROEAVRALAEQADVVLVVGSKNSNSNRLEIAVOL 238  
 QY 241 GKPAYLIDTYODLKQDMLGIEVWGTAGASAPVLYQVBIQDOKAMGSETTSVRENSGI 300  
 DB 239 GKBAFLIDDAKDQGEVWVKVCKCVCVAGASAPDILVQVYVAKQUGGSEALPLE--GR 296  
 QY 301 EKVYVPSIPKELK 313  
 DB 297 ENIVVEVPEKELR 309

## RESULT 7

ISPH\_VIBPA STANDARD; PRT; 321 AA.  
 ID ISPH\_VIBPA  
 AC Q87887;  
 DT 15-SEP-2003 (Rel. 42, Created)  
 DT 15-SEP-2003 (Rel. 42, Last sequence update)  
 DT 15-SEP-2003 (Rel. 42, Last annotation update)  
 DE IspH protein.  
 GN IspH OR LYTB OR VP0537.  
 OS Vibrio parahaemolyticus.  
 CC Bacteria; Proteobacteria; Gammaproteobacteria; Vibrionales;  
 CC Vibrionaceae; Vibrio.  
 CX NCBI\_TaxID=670;  
 RX NCBI\_TaxID=670;

RP SEQUENCE FROM N.A.  
 RC STRAIN=RIMD 2210633 / Serotype O3:K6;  
 RX MEDLINE=22508454; PubMed=12620739;  
 RA Makino K., Oshima K., Kurokawa K., Yokoyama K., Uda T., Tagomori K.,  
 RA Iijima Y., Nishimura M., Nakano M., Yamashita A., Kubota Y., Kimura S.,  
 RA Yasunaga T., Honda T., Shinagawa H., Hattori M., Iida T.;  
 RT "Genome sequence of Vibrio parahaemolyticus: a pathogenic mechanism  
 RT distinct from that of V. Cholerae.";  
 RL Lancet 361:743-749(2003).  
 CC -1- FUNCTION: Converts 1-hydroxy-2-methyl-2-(E)-butenyl 4-diphosphate  
 into isopentenyl diphosphate (IPP) and dimethylallyl diphosphate  
 (DMAPP) (By similarity).  
 CC -1- PATHWAY: Nonmevalonate terpenoid biosynthesis pathway; seventh  
 CC (last) step.  
 CC -1- SIMILARITY: BELONGS TO THE ISPH FAMILY.  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
 CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
 CC EMBL; AP005074; BACS8800.1;  
 DR HAMAP; MF\_00191; -; 1.  
 DR Isoprene biosynthesis; Complete proteome.  
 KW SEQUENCE 321 AA; 35721 MW; DEB4ABA39970798 CRC64;

Query Match 61.0%; Score 989.5; DB 1; Length 321;  
 Best Local Similarity 61.9%; Pred. No. 4.5e-63;  
 Matches 195; Conservative 54; Mismatches 61; Indels 5; Gaps 3;

QY 1 MOYLANPFGFCAGVDALIEYDQALFAFAPITYRHHVHNRTVYDGLKQKGAFFIEEL 60  
 DB 5 MOYLANPFGFCAGVDALIEYDQALFAFAPITYRHHVHNRTVYDGLKQKGAFFIEEL 64  
 QY 61 SDVAVGSYLLIFSAHGVSKVQAEERQITVPDATCPLVTKVHMQVAKHAKGSEVILIG 120  
 DB 65 HEVVDNNITVIFSAHGVSKVQAEERQITVPDATCPLVTKVHMQVAKHAKGSEVILIG 124  
 QY 121 HAGHPEVGTMGQYKCTBGSGIYIVETPEDVRLKVNPNDAVYOTTLISMTDTKYV 178  
 DB 125 HAGHPEVGTMGQYS--SEFGWYLVETPADVRLKAIIVDPSDLFYSQTLISVDETAD 182

```

QY 179 MDALREOPPSIKKQKODICATONRODAVHDLAKISDLITLVGSPNSNSNRLREIAY 238
Db 183 VIDELERRPFPEIQGRKODICATONRODAVHDLAKISDLITLVGSPNSNSNRLREIAY 242
QY 239 QUGKPAVYLIDTYODLKODMLGIEVVGATAGASAEVIVQEVLDOLKAMGETTSVRNS 298
Db 243 KIGTFAVYLIDTYODLKODMLGIEVVGATAGASAEVIVQEVLDOLKAMGETTSVRNS 301
QY 299 GIEKRVFSPKPELK 313
Db 302 GREENPFEPVPELQ 316

```

## RESULT 8

```

ISPH_VIBU
ID ISPH_VIBU STANDARD; PRT; 320 AA.
AC QGDET0;
DT 15-SEP-2003 (Rel. 42, Created)
DT 15-SEP-2003 (Rel. 42, Last sequence update)
DT 15-SEP-2003 (Rel. 42, Last annotation update)
DE ISPH protein.
GN ISPH OR LYTB OR VVI0504.
OS Vibrio vulnificus.
OC Bacteria; Proteobacteria; Gammaproteobacteria; Vibrionales;
OC Vibrionaceae; Vibrio.
OX NCBI_TaxID=672;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=CMCP6;
RA Rhee J.H., Kim S.Y., Chung S.S., Kim J.J., Moon Y.H., Jeong H.,
RA Choy H.B.;
RT "Complete genome sequence of Vibrio vulnificus CMCP6."
RL Submitted (DEC-2002) to the EMBL/GenBank/DBJ databases.
CC -1- FUNCTION: Converts 1-hydroxy-2-methyl-2-(E)-butenyl 4-diphosphate
CC into isopentenyl diphosphate (IPP) and dimethylallyl diphosphate
CC (DMAPP) (By similarity).
CC -1- PATHWAY: Nonmevalonate terpenoid biosynthesis pathway; seventh
CC (last) step.
CC -1- SIMILARITY: BELONGS TO THE ISPH FAMILY.
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@sb-sib.ch).
CC
CC EMBL: AB016798; AAC09023.1;
DR HAMAP; MF_00191; 1.
DR InterPro; IPR003451; LYCB.
DR Pfam; PF02401; LYTB; 1.
DR TIGRFAMs; TIGR00216; isph lytb; 1.
DR Isoprene biosynthesis; Complete proteome.
SQ SEQUENCE 320 AA; 35575 MW; 09F8454ACDA131D8 CRC64;

```

Query Match 60.8%; Score 989; DB 1; Length 320;

Best Local Similarity 62.2%; Pred. No. 4,8e-63;

Matches 196; Conservative 52; Mismatches 61; Indels 6; Gaps 3;

```

QY 1 MOVIANRGFCAGVRAIEIVDAIEAFGAPIYVREHVNRVTVDGLKQKGAFTIEL 60
Db 5 MKILIANRGFCAGVRAISIVRAIEIVQPIYVREHVNRVTVDGLKQKGAFTIEL 64
QY 61 SDVPVGSYLIFSAHGVSKVQQAERQULTVPDAPLTVKVMQVAKAKQGREVILIG 120
Db 65 HEVPDNIIVIFSAHGVQAVRQAKERALTVPDAPLTVKVMQVAKAKQGREVILIG 124
QY 121 HAHGPVBEVGGQYKCTEGGQYIVETPEDVANK--VNNPDILAVYVQTLISMTDTAV 178
Db 125 HAHGPVBEVGGQYKCTEGGQYIVETPEDVANK--VNNPDILAVYVQTLISMTDTAD 182

```

```

QY 179 MDALREOPPSIKKQKODICATONRODAVHDLAKISDLITLVGSPNSNSNRLREIAY 238
Db 183 VIDELERRPFPEIQGRKODICATONRODAVHDLAKISDLITLVGSPNSNSNRLREIAY 242
QY 239 QUGKPAVYLIDTYODLKODMLGIEVVGATAGASAEVIVQEVLDOLKAMGETTSVRNS 298
Db 243 KIGTFAVYLIDTYODLKODMLGIEVVGATAGASAEVIVQEVLDOLKAMGETTSVRNS 300
QY 299 GIEKRVFSPKPELK 313
Db 301 GREENPFEPVPELQ 315

```

## RESULT 9

```

ISPH_VIBU
ID ISPH_VIBU STANDARD; PRT; 316 AA.
AC QGKUA4;
DT 28-FEB-2003 (Rel. 41, Created)
DT 28-FEB-2003 (Rel. 41, Last sequence update)
DT 28-FEB-2003 (Rel. 41, Last annotation update)
DE ISPH protein.
GN ISPH OR LYTB OR VC0685.
OS Vibrio cholerae.
OC Bacteria; Proteobacteria; Gammaproteobacteria; Vibrionales;
OC Vibrionaceae; Vibrio.
OX NCBI_TaxID=666;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=El Tor N16961 / Serotype O1;
RX MEDLINE=20406833; PubMed=10952301;
RA Heidelberg J.F., Eisen J.A., Nelson W.C., Clayton R.A., Gwin M.L.,
RA Dodson R.J., Haft D.H., Hickey E.K., Peterson J.D., Unayen L.A.,
RA Gill S.R., Nelson K.E., Read T.D., Tetstelin H., Richardson D.,
RA Rasmussen M.D., Vamathevan J., Baas S., Qin H., Dragoi I., Sellers P.,
RA McDonald L., Uetzelback T., Fleischmann R.D., Mierman M.C., White O.,
RA Salzberg S.L., Smith H.O., Colwell R.R., Mekalanos J.J., Venter J.C.,
RA Fraser C.M.;
RT "DNA sequence of both chromosomes of the cholera pathogen Vibrio
RT cholerae."
RL Nature 406:477-483 (2000).
CC -1- FUNCTION: Converts 1-hydroxy-2-methyl-2-(E)-butenyl 4-diphosphate
CC into isopentenyl diphosphate (IPP) and dimethylallyl diphosphate
CC (DMAPP) (By similarity).
CC -1- PATHWAY: Nonmevalonate terpenoid biosynthesis pathway; seventh
CC (last) step.
CC -1- SIMILARITY: BELONGS TO THE ISPH FAMILY.
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@sb-sib.ch).
CC
CC EMBL: AE004154; AAF93850.1; ALT_INIT.
DR HAMAP; MF_00191; 1.
DR InterPro; IPR003451; LYCB.
DR Pfam; PF02401; LYTB; 1.
DR TIGRFAMs; TIGR00216; isph lytb; 1.
DR Isoprene biosynthesis; Complete proteome.
SQ SEQUENCE 316 AA; 35156 MW; BB309DBA08C6F79E CRC64;

```

Query Match 60.8%; Score 987; DB 1; Length 316;

Best Local Similarity 61.9%; Pred. No. 6,6e-63;

Matches 195; Conservative 56; Mismatches 58; Indels 6; Gaps 3;

```

QY 1 MOVIANRGFCAGVRAIEIVDAIEAFGAPIYVREHVNRVTVDGLKQKGAFTIEL 60
Db 1 MKILIANRGFCAGVRAISIVRAIEIVQPIYVREHVNRVTVDGLKQKGAFTIEL 60
QY 61 SDVPVGSYLIFSAHGVSKVQQAERQULTVPDAPLTVKVMQVAKAKQGREVILIG 120

```

```

Db 61 NEVPDDNIVIFSAAGVSGAQRQAEKRSITVFDACTCLTVKHEVARBARKEHEVLLIG 120
QY 121 HAGPEVEGTMGQYKTEGGGILVETPEVRNL--KNNPNDAVYVQTLLSMITDKV 178
Db 121 HAGPEVEGTMGQYKTEGGGILVETPEVRNL--KNNPNDAVYVQTLLSMITDKV 178
QY 179 MDALREQPSIKKQKDDICATONRODAVHDLAKISDILVGVSPSSNSNRLEIRIAV 238
Db 179 VIDELRAVFPKIQGPRKDDICATONRODAVHDLAKISDILVGVSPSSNSNRLEIRIAV 238
QY 239 QLGKPAVLIITDYQDKDMLBGLIEVGVITGASAPVLYGVQVITDOLKAMGSETTSVRENS 298
Db 239 KLGTPGLIIPCQPIDPEFVDAQILGVITAGASAPVLYGVQVITDOLKAMGSETTSVRENS 298
QY 299 GIEKVPFISIPKELK 313
Db 297 GREENMFEVFKELQ 311

```

## RESULT 10

```

ISPH_XANAC STANDARD; PRT; 316 AA.
AC 08PNT;
DT 28-FEB-2003 (Rel. 41, Created)
DT 28-FEB-2003 (Rel. 41, Last sequence update)
DE ISPH protein.
ISPH OR LYTB OR XAC1256.
OS Xanthomonas axonopodis (pv. citri).
OC Bacteria; Proteobacteria; Gammaproteobacteria; Xanthomonadales;
Xanthomonadaceae; Xanthomonas.
NCBI_TaxID=92829;
[1]

```

```

RP SEQUENCE FROM N.A.
RC STRAIN=306 / ATCC 13902 / XV 101;
RX MEDLINE=22022145; PubMed=12024217;
da Silva A.C.R., Ferro J.A., Reinach F.C., Farah C.S., Furlan L.R.,
Quaggio R.B., Monteiro-Vitorello C.B., Van Sluys M.A., Almeida N.F.,
Alves L.M.C., do Amaral A.M., Bertolini M.C., Camargo L.B.A.,
Camarote G., Camanho F., Cardoso J., Chambergro F., Chapina L.P.,
Cicarelli R.M.B., Coutinho L.L., Cursino-Santos J.R., El-Dorri H.,
Faria J.B., Ferreira A.J.S., Ferreira R.C.C., Ferro M.I.T.,
Formighieri E.F., Franco M.C., Greggio C.C., Gruber A.,
Katsuyama A.M., Kishi L.T., Leite R.P., Lemos E.G.M., Lemos M.V.F.,
Locali E.C., Machado M.A., Madeira A.M.B.N., Martinez-Rossi N.M.,
Martins E.C., Meidanis J., Menck C.F.M., Miyaki C.Y., Moon D.H.,
Moreira L.M., Novo M.T.M., Okura V.K., Oliveira M.C., Oliveira V.R.,
Pereira H.A., Rossi A., Sena J.A.D., Silva C., de Souza R.F.,
Spindola L.A.F., Takita M.A., Tamura R.E., Teixeira E.C., Tezza R.I.D.,
Trindade dos Santos M., Truffi D., Tsai S.M., White F.F.,
Setubal J.C., Kitajima J.P.;
RT "Comparison of the genomes of two Xanthomonas pathogens with differing
host specificities."
RL Nature 417:459-463(2002).
-1- FUNCTION: Converts 1-hydroxy-2-methyl-2-(E)-butenyl 4-diphosphate
into isopentenyl diphosphate (IPP) and dimethylallyl diphosphate
(DMAPP) (by similarity).
-1- PATHWAY: Nonmevalonate terpenoid biosynthesis pathway; seventh
(last) step.
-1- SIMILARITY: BELONGS TO THE ISPH FAMILY.
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation-
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC EMBL: A011755; AAMB6128.1;
CC HAMAP: MF_00191;
CC InterPro: IPR003451; LytB.

```

```

DR Pfam: PF02401; LytB; 1.
DR Trigram: TRIGR00216; isph_lytB; 1.
KM isoprene biosynthesis; Complete proteome.
SQ SEQUENCE 316 AA; 34690 KM; C931256D6A53D4DA CRC64;

```

Query Match 60.4%; Score 980; DB 1; Length 316;

Best local similarity 61.3%; Pred. No. 2.1e-62;

Matches 192; Conservative 50; Mismatches 69; Indels 2; Gaps 1;

```

QY 1 MCVLANPFGFCAGVDRAIEYDQAIKFAFPIYHHEVNRFTYDGLKQGAVFIEEL 60
Db 1 MCVLANPFGFCAGVDRAIEYDQAIKFAFPIYHHEVNRFTYDGLKQGAVFIEEL 60
QY 61 SDVPVGSYVIFSAHGSXKVOQAESEKOLTFDCACTCLTVKVMQAKAKQREVLLIG 120
Db 61 SDVPVGSYVIFSAHGSXKVOQAESEKOLTFDCACTCLTVKVMQAKAKQREVLLIG 120
QY 61 DEVPDPAITVIFSAHGSQAVROAEKRGKVDACPLVTKVHFVARKCRGRDVLVIG 120
Db 61 DEVPDPAITVIFSAHGSQAVROAEKRGKVDACPLVTKVHFVARKCRGRDVLVIG 120
QY 121 HAGPEVEGTMGQYKTEGGGILVETPEVRNL--KNNPNDAVYVQTLLSMITDKV 180
Db 121 HAGPEVEGTMGQYKTEGGGILVETPEVRNL--KNNPNDAVYVQTLLSMITDKV 180
QY 121 HAGPEVEGTMGQYKTEGGGILVETPEVRNL--KNNPNDAVYVQTLLSMITDKV 180
Db 121 HAGPEVEGTMGQYKTEGGGILVETPEVRNL--KNNPNDAVYVQTLLSMITDKV 180
QY 181 DALREQPSIKKQKDDICATONRODAVHDLAKISDILVGVSPSSNSNRLEIRIAV 240
Db 181 DALREQPSIKKQKDDICATONRODAVHDLAKISDILVGVSPSSNSNRLEIRIAV 240
QY 181 EARARATYPMQGPREDICATONRODAVHDLAKISDILVGVSPSSNSNRLEIRIAV 240
Db 181 EARARATYPMQGPREDICATONRODAVHDLAKISDILVGVSPSSNSNRLEIRIAV 240
QY 241 GXPAYLIDTYQDKDMLBGLIEVGVITGASAPVLYGVQVITDOLKAMGSETTSVRENS 300
Db 241 GXPAYLIDTYQDKDMLBGLIEVGVITGASAPVLYGVQVITDOLKAMGSETTSVRENS 300
QY 241 GXPAYLIDTYQDKDMLBGLIEVGVITGASAPVLYGVQVITDOLKAMGSETTSVRENS 300
Db 241 GXPAYLIDTYQDKDMLBGLIEVGVITGASAPVLYGVQVITDOLKAMGSETTSVRENS 300
QY 301 EKVVPFISIPKELK 313
Db 299 GIEKVPFISIPKELK 311

```

## RESULT 11

```

ISPH_XANAC STANDARD; PRT; 316 AA.
AC 08PBG4;
DT 28-FEB-2003 (Rel. 41, Created)
DT 28-FEB-2003 (Rel. 41, Last sequence update)
DE ISPH protein.
ISPH OR LYTB OR XCC1157.
OS Xanthomonas campestris (pv. campestris).
OC Bacteria; Proteobacteria; Gammaproteobacteria; Xanthomonadales;
Xanthomonadaceae; Xanthomonas.
NCBI_TaxID=340;
[1]

```

```

RP SEQUENCE FROM N.A.
RC STRAIN=ATCC 33913 / NCPPB 528;
RX MEDLINE=22022145; PubMed=12024217;
da Silva A.C.R., Ferro J.A., Reinach F.C., Farah C.S., Furlan L.R.,
Quaggio R.B., Monteiro-Vitorello C.B., Van Sluys M.A., Almeida N.F.,
Alves L.M.C., do Amaral A.M., Bertolini M.C., Camargo L.B.A.,
Camarote G., Camanho F., Cardoso J., Chambergro F., Chapina L.P.,
Cicarelli R.M.B., Coutinho L.L., Cursino-Santos J.R., El-Dorri H.,
Faria J.B., Ferreira A.J.S., Ferreira R.C.C., Ferro M.I.T.,
Formighieri E.F., Franco M.C., Greggio C.C., Gruber A.,
Katsuyama A.M., Kishi L.T., Leite R.P., Lemos E.G.M., Lemos M.V.F.,
Locali E.C., Machado M.A., Madeira A.M.B.N., Martinez-Rossi N.M.,
Martins E.C., Meidanis J., Menck C.F.M., Miyaki C.Y., Moon D.H.,
Moreira L.M., Novo M.T.M., Okura V.K., Oliveira M.C., Oliveira V.R.,
Pereira H.A., Rossi A., Sena J.A.D., Silva C., de Souza R.F.,
Spindola L.A.F., Takita M.A., Tamura R.E., Teixeira E.C., Tezza R.I.D.,
Trindade dos Santos M., Truffi D., Tsai S.M., White F.F.,
Setubal J.C., Kitajima J.P.;
RT "Comparison of the genomes of two Xanthomonas pathogens with differing
host specificities."
RL Nature 417:459-463(2002).
-1- FUNCTION: Converts 1-hydroxy-2-methyl-2-(E)-butenyl 4-diphosphate
into isopentenyl diphosphate (IPP) and dimethylallyl diphosphate
(DMAPP) (by similarity).
-1- PATHWAY: Nonmevalonate terpenoid biosynthesis pathway; seventh
(last) step.
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation-
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC EMBL: A011755; AAMB6128.1;
CC HAMAP: MF_00191;
CC InterPro: IPR003451; LytB.

```

(last) step.  
 CC -1- SIMILARITY: BELONGS TO THE ISPH FAMILY.  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
 CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
 CC  
 CC EMBL: AB012214; AAM40456.1; -  
 DR HAMAP: ME\_00191; -1.  
 DR InterPro: IPR003451; LYTb.  
 DR Pfam: PF02401; LYTb; 1.  
 DR TIGRFAMs: TIGR00216; isph\_lytB; 1.  
 DR Isoprene biosynthesis; Complete proteome.  
 KW Isoprene biosynthesis; Complete proteome.  
 SQ SEQUENCE 316 AA; 34641 MW; 6BFA272A4CA1EAC CRC64;

Query Match 59.8%; Score 970; DB 1; Length 316;  
 Best local similarity 61.3%; Pred. No. 1.1e-6;  
 Matches 192; Conservative 48; Mismatches 71; Indels 2; Gaps 1;

QY 1 MOVLANPRGFCAGVRAIIVQDAIEAFGAPIYVHEVHNRTVVDGKQGAFTIEEL 60  
 DB 1 MDVLANPRGFCAGVRAIIVQDAIEAFGAPIYVHEVHNRTVVDGKQGAFTIEEL 60  
 QY 61 SDVPGSYLIFSAHVSKEVQQAEROLTPDAPCPVTKVMQVAKAKQGREVILIG 120  
 DB 61 DEVPDATVIFSAHVSQAVRVEERGLKVPDAPCPVTKVMQVAKAKQGREVILIG 120  
 QY 121 HAHGPEVEGTMQYEKTEGGGIIYVTEPEDVRLKVNPNDAVYVTTLSMTDKVMV 180  
 DB 121 HAHGPEVEGTMQYSEKRGPGQIYVTEPEDVRLKVNPNDAVYVTTLSMTDKVMV 180  
 QY 121 HAHGPEVEGTMQYSEKRGPGQIYVTEPEDVRLKVNPNDAVYVTTLSMTDKVMV 180  
 DB 121 HAHGPEVEGTMQYSEKRGPGQIYVTEPEDVRLKVNPNDAVYVTTLSMTDKVMV 180  
 QY 181 DALRQPSIKQKQKDDICATONRODAVHDLAKISDLILVYVSSNSNRRLRIAYOL 240  
 DB 181 DALRQPSIKQKQKDDICATONRODAVHDLAKISDLILVYVSSNSNRRLRIAYOL 240  
 QY 181 EALKRYPPAQGRKHDDICATONRODAVHDLAKISDLILVYVSSNSNRRLRIAYOL 240  
 DB 181 EALKRYPPAQGRKHDDICATONRODAVHDLAKISDLILVYVSSNSNRRLRIAYOL 240  
 QY 241 GKPAYLIDTYODLKQDMLEIEVGVTAGASAPVYLVOEVIDOLKAMGETTSVRENSGI 300  
 DB 241 GVSYSYLDNNSRIDPMIYVQKHIGTAGASAPVYLVOEVIDOLKAMGETTSVRENSGI 300  
 QY 301 EEKVVSIPKEIK 313  
 DB 299 PESWFFALPEIKR 311

RESULT 12  
 ISPH\_HABIN STANDARD; PRT; 314 AA.  
 ID ISPH\_HABIN  
 AC P44976;  
 DT 01-NOV-1995 (Rel. 32, Created)  
 DT 01-NOV-1995 (Rel. 32, Last sequence update)  
 DT 28-FEB-2003 (Rel. 41, Last annotation update)  
 DE ISPH protease.  
 GN ISPH OR LYTb OR H11007.  
 OS Haemophilus influenzae.  
 OC Bacteria; Proteobacteria; Gammaproteobacteria; Pasteurellales;  
 CC Pasteurellaceae; Haemophilus.  
 NCBI\_TaxID=727;  
 RX MEDLINE=95350630; PubMed=7542800;  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=Rd / KM20 / ATCC 51907;  
 RA Fleischmann R.D., Adams M.D., White O., Clayton R.A., Kirkness E.F.,  
 Kerlavage A.R., Bult C.J., Tomb J.-F., Dougherty B.A., Merrick J.M.,  
 McKenney K., Sutton G., Fitzhugh W., Fields C.A., Gocayne J.D.,  
 Scott J.D., Shiley R., Liu L.-I., Glodek A., Kelley J.M.,  
 Weidman J.F., Phillips C.A., Spriggs T., Hedlorn E., Cotton M.D.,  
 Usterback T.R., Hanna M.C., Nguyen D.T., Sandek D.M., Brandon R.C.,  
 R.A. Fine L.D., Fritchman J.L., Fuhrmann J.L., Geoghagen N.S.M.,  
 Gnehm C.L., McDonald L.A., Small K.V., Fraser C.K., Smith H.O.,

RA Venter J.C.;  
 RT "Whole-genome random sequencing and assembly of Haemophilus influenzae  
 Rd".  
 RL Science 269:496-512(1995).  
 CC -1- FUNCTION: Converts 1-hydroxy-2-methyl-2-(E)-butenyl 4-diphosphate  
 CC into isopentenyl diphosphate (IPP) and dimethylallyl diphosphate  
 CC (DMAPP) (By similarity).  
 CC -1- PATHWAY: Nonmevalonate terpenoid biosynthesis pathway; seventh  
 CC (last) step.  
 CC -1- SIMILARITY: BELONGS TO THE ISPH FAMILY.  
 CC  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
 CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
 CC  
 CC EMBL: U32781; AAC22668.1; -  
 DR TIGR: H11007; -1.  
 DR HAMAP: ME\_00191; -1.  
 DR InterPro: IPR003451; LYTb.  
 DR Pfam: PF02401; LYTb; 1.  
 DR TIGRFAMs: TIGR00216; isph\_lytB; 1.  
 DR Isoprene biosynthesis; Complete proteome.  
 KW Isoprene biosynthesis; Complete proteome.  
 SQ SEQUENCE 314 AA; 34468 MW; 44860D11DE9F9EBC CRC64;

Query Match 59.7%; Score 969; DB 1; Length 314;  
 Best local similarity 58.8%; Pred. No. 1.2e-61;  
 Matches 184; Conservative 61; Mismatches 64; Indels 4; Gaps 2;

QY 1 MOVLANPRGFCAGVRAIIVQDAIEAFGAPIYVHEVHNRTVVDGKQGAFTIEEL 60  
 DB 1 MDVLANPRGFCAGVRAIIVQDAIEAFGAPIYVHEVHNRTVVDGKQGAFTIEEL 60  
 QY 61 SDVPGSYLIFSAHVSKEVQQAEROLTPDAPCPVTKVMQVAKAKQGREVILIG 120  
 DB 61 DEVPDATVIFSAHVSQAVRVEERGLKVPDAPCPVTKVMQVAKAKQGREVILIG 120  
 QY 121 HAHGPEVEGTMQYEKTEGGGIIYVTEPEDVRLKVNPNDAVYVTTLSMTDKVMV 180  
 DB 121 HAHGPEVEGTMQYSEKRGPGQIYVTEPEDVRLKVNPNDAVYVTTLSMTDKVMV 180  
 QY 121 HAHGPEVEGTMQYSEKRGPGQIYVTEPEDVRLKVNPNDAVYVTTLSMTDKVMV 180  
 DB 121 HAHGPEVEGTMQYSEKRGPGQIYVTEPEDVRLKVNPNDAVYVTTLSMTDKVMV 180  
 QY 181 DALRQPSIKQKQKDDICATONRODAVHDLAKISDLILVYVSSNSNRRLRIAYOL 240  
 DB 181 DALRQPSIKQKQKDDICATONRODAVHDLAKISDLILVYVSSNSNRRLRIAYOL 240  
 QY 179 EALKRYPPAQGRKHDDICATONRODAVHDLAKISDLILVYVSSNSNRRLRIAYOL 240  
 DB 179 EALKRYPPAQGRKHDDICATONRODAVHDLAKISDLILVYVSSNSNRRLRIAYOL 240  
 QY 241 GKPAYLIDTYODLKQDMLEIEVGVTAGASAPVYLVOEVIDOLKAMGETTSVRENSGI 300  
 DB 239 GYSQGLDBPADIDQDMVNDVKITGITAGASAPVYLVOEVIDOLKAMGETTSVRENSGI 300  
 QY 301 EEKVVSIPKEIK 313  
 DB 297 EGNWFFVPEIKR 309

RESULT 13  
 ISPH\_PASMO STANDARD; PRT; 314 AA.  
 ID ISPH\_PASMO  
 AC P57960;  
 DT 16-OCT-2001 (Rel. 40, Created)  
 DT 16-OCT-2001 (Rel. 40, Last sequence update)  
 DT 28-FEB-2003 (Rel. 41, Last annotation update)  
 DE ISPH protein.  
 GN ISPH OR LYTb OR PM1664.  
 OS Pasteurella multocida.  
 OC Bacteria; Proteobacteria; Gammaproteobacteria; Pasteurellales;  
 CC Pasteurellaceae; Pasteurella.  
 NCBI\_TaxID=747;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=Pm70;

```

RX MEDLINE=21145866; PubMed=11248100;
RA May B.J., Zhang Q., Li L., Pautsien M.L., Whittem T.S., Kapur V.;
RT "Complete genomic sequence of Pasteurella multocida Fm70."
RL Proc. Natl. Acad. Sci. U.S.A. 98:3460-3465(2001).
CC
CC -1- FUNCTION: Converts 1-hydroxy-2-methyl-2-(E)-butenyl 4-diphosphate
into isopentenyl diphosphate (IPP) and dimethylallyl diphosphate
(DMAPP) (By similarity).
CC -1- PATHWAY: Nonmevalonate terpenoid biosynthesis pathway; seventh
(last) step.
CC
CC -1- SIMILARITY: BELONGS TO THE ISPP FAMILY.
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
between the Swiss Institute of Bioinformatics and the EMBL Outstation
at the European Bioinformatics Institute. There are no restrictions on its
use by non-profit institutions as long as its content is in no way
modified and this statement is not removed. Usage by and for commercial
entities requires a license agreement (See http://www.isb-sib.ch/announce/
or send an email to license@isb-sib.ch.)
CC
CC EMBL; AB006203; AAK03748.1; -.
DR HAMAP; MF_00191; -.
DR InterPro; IPR003451; LYTB.
DR Pfam; PF02401; LYTB; 1.
DR TIGRfams; TIGR00216; ispp_lytb; 1.
DR isoprene biosynthesis; Complete proteome.
SQ SEQUENCE 314 AA; 34469 MM; DCB6179787883125 CRC64;

Query Match 59.7%; Score 969; DB 1; Length 314;
Best Local Similarity 59.1%; Pred. No. 1.2e-61;
Matches 185; Conservative 60; Mismatches 64; Indels 4; Gaps 2;

QY 1 MQVLNPRGFCAGVDRAIBIVDAIBAFGAPIVYRHEVNNRTVVDLKKQGAFFIEL 60
DB 1 MKLILNPRGFCAGVDRAISIVSALRIHSGPIYVRHEVNNRTVVDLKKQGAFFIEL 60
QY 61 SDVWGSYLIFSAGVSKVQGEAEERQLVFDATCPLVTKVHQAQAKAGREVLIG 120
DB 61 HEVVDGAIIVFSAGVSAQVAREKKNLKFVDTCPVTKVHQAQAKAGREVLIG 120
QY 121 HAHEVEGTMGQYEKCTEGGIVLVETPEDVRNLKVNPNDAVYVQTLLSMDFKWM 180
DB 121 HEHREVEGTMGQYD--NPEGGIVLVSVEDIALGRLQEDDLTFMTQTLLSDTFSEVI 178
QY 181 DALREPPSTIREKQKDDICTATONRQDAVHDLAKISDLILVSGSPNSNSNRLAEIAVOL 240
DB 179 EALKQKTPALQGPXKNDICVATTRQQAVERLAQOSDLIVVSGKSNNSNRLAEIASRM 238
QY 241 GKPAVLIDTYODLKGDLGIEVVGVTAGASAPENVVQEVITDOLKAMGGTTSVRENSGI 300
DB 239 GVVSKLIDDANDIDPTMLEVNTTIGTAGASAPENVVQSVKRLQELG--VTRVBELOGC 296
QY 301 EEKVVFSIPKELK 313
DB 297 EKVTFEVPRELR 309

RESULT 14
ISPP_NEIMA STANDARD; PRT; 322 AA.
AC Q9JH39;
DT 28-FEB-2003 (Rel. 41, Last sequence update)
DT 28-FEB-2003 (Rel. 41, Last annotation update)
DE ISPP protein.
OS ISPP OR LYTB OR NMA0624 OR NMB1831.
OS Neisseria meningitidis (serogroup A), and
OC Neisseria meningitidis (serogroup B).
OC Bacteria; Proteobacteria; Betaproteobacteria; Neisseriales;
OC Neisseriaceae; Neisseria.
OX NCBI_TaxID=56599, 491;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=Z2491 / Serogroup A / Serotype 4A;

```

```

RX MEDLINE=2022556; PubMed=10761919;
RA Parthill J., Achtman M., James K.D., Bentley S.D., Churcher C.,
RA Klee S.R., Morelli G., Basham D., Brown D., Chillingworth T.,
RA Davies R.M., Davis P., Devlin K., Felwell T., Hamlin N., Holroyd S.,
RA Jagsels K., Leather S., Moule S., Mungall K., Quail M.A.,
RA Rajadream M.A., Rutherford K.M., Simmonds M., Skelton J.,
RA Whitehead S., Spratt B.G., Barrall B.G.;
RT "Complete DNA sequence of a serogroup A strain of Neisseria
meningitidis Z2491."
RL Nucleic Acids Res. 28(1):502-506(2000).
RN [2]
RP SEQUENCE FROM N.A.
RC STRAIN=MC58 / Serogroup B;
RX MEDLINE=20175755; PubMed=10710307;
RA Tettein H., Saunders N.J., Heideberg J., Jeffries A.C., Nelson K.E.,
RA Eisen J.A., Ketchum K.A., Hood D.W., Peden J.F., Dodson R.J.,
RA Nelson W.C., Gwyn M.L., Deboy R., Peterson J.D., Hickey E.K.,
RA Haft D.H., Salzberg S.L., White O., Fleischmann R.D., Dougherty B.A.,
RA Mason T., Ciecko A., Parksey D.S., Blair E., Citron H., Clark E.B.,
RA Cotton M.D., Uetereback T.R., Khouri H., Qin H., Vamathevan J.,
RA Gill J., Scarlato V., Maignani V., Pizsa M., Grandi G., Sun L.,
RA Smith H.O., Fraser C.M., Koxon E.R., Rappelli R., Venter J.C.;
RT "Complete genome sequence of Neisseria meningitidis serogroup B strain
MC58."
RL Science 287:1809-1815(2000).
CC -1- FUNCTION: Converts 1-hydroxy-2-methyl-2-(E)-butenyl 4-diphosphate
into isopentenyl diphosphate (IPP) and dimethylallyl diphosphate
(DMAPP) (By similarity).
CC -1- PATHWAY: Nonmevalonate terpenoid biosynthesis pathway; seventh
(last) step.
CC
CC -1- SIMILARITY: BELONGS TO THE ISPP FAMILY.
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
between the Swiss Institute of Bioinformatics and the EMBL Outstation
at the European Bioinformatics Institute. There are no restrictions on its
use by non-profit institutions as long as its content is in no way
modified and this statement is not removed. Usage by and for commercial
entities requires a license agreement (See http://www.isb-sib.ch/announce/
or send an email to license@isb-sib.ch.)
CC
CC EMBL; AL162753; CAB83914.1; -.
DR EMBL; AB002532; AAF42166.1; -.
DR PIR; D81038; D81038.
DR TIGR; NMB1831; -.
DR HAMAP; MF_00191; -.
DR InterPro; IPR003451; LYTB.
DR Pfam; PF02401; LYTB; 1.
DR TIGRfams; TIGR00216; ispp_lytb; 1.
DR isoprene biosynthesis; Complete proteome.
SQ SEQUENCE 322 AA; 35322 MM; 6B92762C0653D78F CRC64;

Query Match 59.6%; Score 968; DB 1; Length 322;
Best Local Similarity 61.5%; Pred. No. 1.5e-61;
Matches 192; Conservative 48; Mismatches 66; Indels 6; Gaps 2;

QY 3 IVLNPRGFCAGVDRAIBIVDAIBAFGAPIVYRHEVNNRTVVDLKKQGAFFIELSD 62
DB 6 IILNPRGFCAGVDRAISIVSALRIHSGPIYVRHEVNNRTVVDLKKQGAFFIELSD 65
QY 63 VFWGSYLIFSAGVSKVQGEAEERQLVFDATCPLVTKVHQAQAKAGREVLIG 122
DB 66 VEPGATLVISAGVSKVQGEAEERQLVFDATCPLVTKVHQAQAKAGREVLIG 125
QY 123 GHPEVEGTMGQYEKCTEGGIVLVETPEDVRNLKVNPNDAVYVQTLLSMDFKWM 182
DB 126 GHVREVEGTMGQYD-----GKMLVETVGVDAKLEVRNPDKLAVYVQTLTSLVDETQIIIA 181
QY 183 LREPPSTIREKQKDDICTATONRQDAVHDLAKISDLILVSGSPNSNSNRLAEIAVOL 242
DB 182 LNAEPNNRNPKEKDDICVATTRQQAVERLAQOSDLIVVSGKSNNSNRLAEIASRGI 241
QY 243 PAVLIDTYODLKGDLGIEVVGVTAGASAPENVVQEVITDOLKAMGGTTSVRENSGIEE 302
DB 243 PAVLIDTYODLKGDLGIEVVGVTAGASAPENVVQEVITDOLKAMGGTTSVRENSGIEE 302

```

DB 242 DAYWDMAGYIQRAMFEGKXKGVTAQASAPVLYREVLATINQGHRT--VREGBAER 299  
 QY 303 KYVFSIPKELKX 314  
 DB 300 SIVFVLPKELR 311

RESULT 15  
 ISPH\_XYLFA STANDARD; PRT; 316 AA.

AC 092AS9;  
 DT 28-FEB-2003 (Rel. 41, Created)  
 DT 28-FEB-2003 (Rel. 41, Last sequence update)  
 DT 15-SEP-2003 (Rel. 42, Last annotation update)  
 DE ISPH protein.  
 GN ISPH OR LYTB OR XP2416 OR PD1435.  
 OS Xylella fastidiosa (strain Temecula / ATCC 700964).  
 OC Bacteria; Proteobacteria; Gammaproteobacteria; Xanthomonadales;  
 OC Xanthomonadaceae; Xylella.  
 NCBI TaxID=3371, 183190;  
 [1]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=945C;  
 RX MEDLINE=20365717; PubMed=10910347;  
 RA Simpson A.J.G., Reinach F.C., Arruda P., Abreu F.A., Aencio M.,  
 Alvares R., Alves L.M.C., Araya J.E., Baia G.S., Baptista C.S.,  
 Barros M.H., Bonaccorsi E.D., Bordin S., Bove J.M., Birones M.R.S.,  
 Bueno M.R.P., Camargo A.A., Camargo L.E.A., Carro D.M., Carrer H.,  
 Colauto N.B., Colombo C., Costa P.F., Costa M.C.R., Costa-Neto C.M.,  
 Coutinho L.B., Cristofani M., Dias-Neto E., Docena C., El-Dorry H.,  
 Faciniani A.P., Ferreira A.J.S., Ferreira V.C.A., Ferro J.A.,  
 Fraga J.S., Franca S.C., Franco M.C., Frome M., Furian L.R.,  
 Gantier M., Goldman G.H., Goldman M.H.S., Gomes S.L., Gruber A.,  
 Ho P.L., Hobeisel J.D., Junqueira M.L., Kemper E.L., Kitajima J.P.,  
 Krieger J.E., Kurane E.E., Laigret F., Lambis M.R., Leite L.C.C.,  
 Lemos E.G.M., Lemos M.V.F., Lopes S.A., Lopes C.R., Machado J.A.,  
 Machado M.A., Madeira A.M.B.N., Madeira H.M.F., Marino C.L.,  
 Marques M.V., Martins E.A.L., Martins E.M.F., Macskuma A.Y.,  
 Menck C.F.M., Miracca E.C., Miyaki C.Y., Monteiro-Vitorello C.B.,  
 Moon D.H., Nagel M.A., Nascimento A.L.T.O., Netto L.E.S.,  
 RA Niani A.Jr., Nobrega F.G., Nunes L.R., Oliveira M.A.,  
 RA de Oliveira M.C., de Oliveira R.C., Palmieri D.A., Paris A.,  
 RA de Oliveira R.R., Pereira G.A.G., Pereira H.A. Jr., Peguero J.B.,  
 RA Quaggio R.B., Roberto P.G., Rodrigues V., de Rosa A.J.M.,  
 RA de Rosa V.E. Jr., de Sa R.G., Santelli R.V., Sawasaki H.E.,  
 RA da Silva A.C.R., da Silva A.M., da Silva F.R., Silva W.A. Jr.,  
 RA da Silva J.F., Silvestri M.L.Z., Siqueira W.J., de Souza A.A.,  
 RA de Souza A.P., Terenzi M.F., Truffi D., Tsai S.M., Tsubako M.H.,  
 RA Vallada H., Van Sluys M.A., Verjovski-Almeida S., Vettore A.L.,  
 RA Zago M.A., Zatz M., Zeldenis J., Zetubal J.C.,  
 RT "The genome sequence of the plant pathogen Xylella fastidiosa,"  
 RT Nature 406:151-159(2000).  
 [2]  
 R2 SEQUENCE FROM N.A.  
 R3 STRAIN=Temecula; ATCC 700964;  
 RX MEDLINE=22421331; PubMed=12533478;  
 RA Van Sluys M.A., de Oliveira M.C., Monteiro-Vitorello C.B.,  
 RA Miyaki C.Y., Furian L.R., Camargo L.E.A., da Silva A.C.R., Moon D.H.,  
 RA Takita M.A., Lemos E.G.M., Machado M.A., Ferro M.I.T., da Silva F.R.,  
 RA Goldman M.H.S., Goldman G.H., Lemos M.V.F., El-Dorry H., Tsai S.M.,  
 RA Carier H., Carro D.M., de Oliveira R.C., Nunes L.R., Siqueira W.J.,  
 RA Coutinho L.B., Kimura E.T., Ferro E.S., Harakava R., Kurane E.E.,  
 RA Marinho C.L., Gigliotti E., Abreu I.L., Alves L.M.C., do Amaral A.M.,  
 RA Baia G.S., Blanco S.R., Brito M.S., Camargo F.S., Celestino A.V.,  
 RA da Cunha A.F., Fenille R.C., Ferro J.A., Formighieri E.P., Kishi L.T.,  
 RA Leon S.G., Oliveira A.R., Rosa V.E. Jr., Sasaki F.T., Sena J.A.D.,  
 RA de Souza A.A., Truffi D., Tsubako M.H., Tsai S.M., Zago M.A.,  
 RA Civerolo E.L., Simpson A.J.G., Almeida N.P. Jr., Zetubal J.C.,  
 RA Kitajima J.P.;  
 RT "Comparative analyses of the complete genome sequences of Pierce's  
 RT disease and citrus variegated chlorosis strains of Xylella  
 RT fastidiosa,"

RL J. Bacteriol. 185:1018-1026 (2003).  
 CC -1- FUNCTION: Converts 1-hydroxy-2-methyl-2-(E)-butenyl 4-diphosphate  
 CC into isopentenyl diphosphate (IPP) and dimethylallyl diphosphate  
 CC (DMAPP) (By similarity)  
 CC -1- PATHWAY: Nonmevalonate terpenoid biosynthesis pathway; seventh  
 CC (last) step.  
 CC -1- SIMILARITY: BELONGS TO THE ISPH FAMILY.  
 CC  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (see <http://www.isb-sib.ch/announce/>  
 CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
 CC  
 DR EMBL: AB004050; AAF85215.1; -  
 DR EMBL: AB012558; AAC29279.1; -  
 DR PIR: C82561; C82561.  
 DR HAMAP: MF\_00191; -; 1.  
 DR InterPro: IPR003451; LYTB.  
 DR Pfam: PF02401; LYTB; 1.  
 DR TIGRFAMs: TIGR00216; isph\_Lytb; 1.  
 KW Isoprene biosynthesis; Complete proteome.  
 SQ SEQUENCE 316 AA; 34704 MW; 1A2E80B9A98D334A CRC64;

Query Match 59.6%; Score 967; DB 1; Length 316;  
 Best Local Similarity 61.0%; Pred. No. 1.7e-61;  
 Matches 191; Conservative 53; Mismatches 67; Indels 2; Gaps 1;

1 MOYIYANRPGACGADRAIRIYDQALEAFGAPTYRHHVNRHTVVDGKOGAVFIEEL 60  
 1 MYVLANPFGAGVADRAIRIYDQALEAFGAPTYRHHVNRHTVVDGKOGAVFIEEL 60  
 61 SVPVGSYLIIFSAHSGSEKVOCEAEERDITVFDATCPVTXKVMQVAKRAGREVILG 120  
 61 HGVPGATVIFSAHSGSEKVOCEAEERDITVFDATCPVTXKVMQVAKRAGREVILG 120  
 121 HNGHEVBCTMGQYKCTBGGIYIVETPEYRNLKANNPDIAVYOTLSMTQKAV 180  
 121 HNGHEVBCTMGQYKCTBGGIYIVETPEYRNLKANNPDIAVYOTLSMTQKAV 180  
 121 HNGHEVBCTMGQYKCTBGGIYIVETPEYRNLKANNPDIAVYOTLSMTQKAV 180  
 181 DALRQPSIKKQKDDICVATONRQDAVHDIAKISDILLVGSNNSSNLRIRIAVQL 240  
 181 DALRQPSIKKQKDDICVATONRQDAVHDIAKISDILLVGSNNSSNLRIRIAVQL 240  
 241 GRVAVLDTYDQKQDMLGIEVGVTAQASAPVLYREVLATINQGHRT--VREGBAER 300  
 241 GYASVYLDSAMRIDPAWYIDKHIGVTAQASAPVLYREVLATINQGHRT--VREGBAER 300  
 301 EKVFSIPKELK 313  
 299 PSMVFLPKELR 311

Search completed: January 29, 2004, 15:50:43  
 Job time : 6.98823 secs

GenCore version 5.1.6  
Copyright (c) 1993 - 2004 Compugen Ltd.

OM protein - protein search, using sw model

Run on: January 29, 2004, 15:44:14 / Search time 23.7034 Seconds  
(without alignments)  
3461.975 Million cell updates/sec

Title: US-09-941-947a-18

Perfect score: 1623

Sequence: 1 MGVLANPRGFCAGVDRAIE.....GIEKVVPSIPKELKGMQA 318

Scoring table: BLOSUM62

Gapop 10.0, Gapext 0.5

Searched: 830525 seqs, 258052604 residues

Total number of hits satisfying chosen parameters: 830525

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database:

SPTREML\_23:\*

- 1: sp\_archaea:\*
- 2: sp\_bacteria:\*
- 3: sp\_fungi:\*
- 4: sp\_human:\*
- 5: sp\_invertebrate:\*
- 6: sp\_mammal:\*
- 7: sp\_phc:\*
- 8: sp\_organelle:\*
- 9: sp\_phage:\*
- 10: sp\_plant:\*
- 11: sp\_rodent:\*
- 12: sp\_virus:\*
- 13: sp\_vertebrate:\*
- 14: sp\_unclassified:\*
- 15: sp\_virus:\*
- 16: sp\_bacteriophage:\*
- 17: sp\_archaeap:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

## SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	989	60.9	320	16	Q8DET0
2	958	59.0	318	16	Q8EB17
3	866	53.4	315	16	Q8D2R2
4	764	47.1	336	16	Q8FOP0
5	751	46.3	372	16	Q8G2S7
6	658	40.5	352	16	Q8G4I8
7	628	38.7	330	16	Q8G3I3
8	599	36.9	535	5	Q8I295
9	598.5	36.9	535	5	Q8BUX6
10	404.5	24.9	827	16	Q8R152
11	393	22.1	299	16	Q8EMR9
12	358	22.1	642	16	Q8T109
13	343	21.1	402	16	Q8DK29
14	305	18.8	464	10	Q8FEP0
15	300	18.5	466	10	Q84B35
16	265	16.3	332	16	Q8KFN9

17	230	14.2	452	10	Q9SZ00	Q9SZ00 arabidopsis
18	176.5	10.9	167	10	Q9SP19	Q9SP19 nicotiana t
19	170	10.5	221	10	Q94ER5	Q94ER5 brassica ra
20	170	10.5	223	10	Q94ER4	Q94ER4 brassica ra
21	150.5	9.3	130	10	Q8LKH3	Q8LKH3 forsythia l
22	115	7.1	929	16	Q86810	Q86810 streptomyce
23	112.5	6.9	434	16	Q8S5X9	Q8S5X9 thermotoga
24	112	6.9	969	16	Q8G4H3	Q8G4H3 bifidobacte
25	111.5	6.9	670	2	P96626	P96626 bacillus su
26	111.5	6.9	777	16	Q31489	Q31489 bacillus su
27	110	6.8	500	16	Q8YA67	Q8YA67 listeria mo
28	110	6.8	542	2	Q8RR20	Q8RR20 listeria mo
29	109	6.7	499	16	Q92EY8	Q92EY8 listeria in
30	109	6.7	925	5	Q22758	Q22758 caenorhabdi
31	108.5	6.7	596	16	Q97MA7	Q97MA7 clostridium
32	108.5	6.7	602	16	Q8RBX6	Q8RBX6 thermoaer
33	108	6.7	886	16	Q05560	Q05560 mycobacteri
34	107.5	6.6	501	5	Q61118	Q61118 leishmania
35	107.5	6.5	960	11	Q8BSZ1	Q8BSZ1 mus musculu
36	105.5	6.4	527	16	Q98AG2	Q98AG2 rhizobium l
37	104.5	6.4	488	16	Q8CMQ7	Q8CMQ7 staphylococ
38	103.5	6.4	482	15	Q99F12	Q99F12 chimpanzee
39	103.5	6.4	545	16	Q8CX13	Q8CX13 oceanobacill
40	103.5	6.4	960	11	Q8C1E4	Q8C1E4 mus musculu
41	103	6.3	374	16	Q8X8F4	Q8X8F4 escherichia
42	103	6.3	582	5	Q9U5L7	Q9U5L7 paracentrot
43	103	6.3	713	16	Q8U9Z5	Q8U9Z5 agrobacteri
44	102.5	6.3	349	1	Q49598	Q49598 methanopyru
45	102.5	6.3	396	17	Q96Y01	Q96Y01 sulfobobus

## ALIGNMENTS

## RESULT 1

Q8DET0 PRELIMINARY: PRT: 320 AA.

AC Q8DET0;  
DT 01-MAR-2003 (TREMBLrel. 23, Created)  
DT 01-MAR-2003 (TREMBLrel. 23, Last sequence update)  
DT 01-MAR-2003 (TREMBLrel. 23, Last annotation update)  
DE Penicillin tolerance protein.  
GN VY10504.  
OS Vibrio vulnificus.  
OC Bacteria; Proteobacteria; Gammaproteobacteria; Vibrionales;  
OC Vibrionaceae; Vibrrio.  
OX NCBI\_Taxid=672;  
RN [1]  
RP SOURCE FROM N.A.  
RC STRAIN=CMCP6;  
RA Rhee J.H., Kim S.Y., Chung S.S., Kim J.J., Moon Y.H., Jeong H.,  
RA Choy H.E.;  
RT "Complete genome sequence of Vibrio vulnificus CMCP6."  
RL Submitted (DEC-2002) to the EMBL/Genbank/DBJ databases.  
DR EMBL; AE016798; AAO09023.1; -  
KW Complete proteome.  
SQ SEQUENCE 320 AA; 35575 MW; 097B45ACDA1B1D8 CRC64;

Query Match 60.9%; Score 989; DB 16; Length 320;

Best Local Similarity 62.2%; Pred. No. 2.7e-66;

Matches 196; Conservative 52; Mismatches 61; Indels 6; Gaps 3;

QY	1	MGVLANPRGFCAGVDRAIEYDQALTEAFCAPIYVRRHRTVWDGKQGAVIDEL	60
QY	1		60
QY	61	SDVPVGSYLIFSAGVSKRVQDAERQTVFDPATCPVTYVHMVAKAKQGREVILIG	120
QY	61		120
QY	62	HEVPDQNIYIFSAHGVSGVROEAKERALTVDATCPVTYVHMVAKAKQGREVILIG	124
QY	62		124
QY	121	HAGHEVEGTQVQKCTBGGGIYVVEPVEDYRNK--VNNNDLAVYQTTLSMTDTKV	178
QY	121		178
QY	122	HAGHEVEGTQVQKCTBGGGIYVVEPVEDYRNK--VNNNDLAVYQTTLSMTDTKV	178
QY	122		178
QY	123	HAGHEVEGTQVQKCTBGGGIYVVEPVEDYRNK--VNNNDLAVYQTTLSMTDTKV	182
QY	123		182

QY 179 MYDALREQPSIKEOKKDDICATATONRODAVHDLAKISDLILVYSGSPSSNSNRLRELAIV 238  
 DB 183 VIDELARRVFPETIGRKDDICATATONRODAVRELAKDDVYVYSGSKSSNSTLKEIAE 242  
 QY 239 QLGKPAVILIDYODIKODMLBGIEVGVTAGASAEVLVQVVDQLKAMGGETTSVRENS 298  
 DB 243 KLGTFAVILIDYODIEPQMFQKGVGVTAGASAEVLVQVVDQLKAMGGETTSVRENS 300  
 QY 299 GIEKVFSPKPELK 313  
 DB 301 GREENMFPEVPELQ 315

## RESULT 2

Q8EB17 PRELIMINARY; PRT; 318 AA.  
 ID 08EB17  
 AC 08EB17  
 DT 01-MAR-2003 (Tremblrel. 23, Created)  
 DT 01-MAR-2003 (Tremblrel. 23, Last sequence update)  
 DT 01-MAR-2003 (Tremblrel. 23, Last annotation update)  
 DE Penicillin tolerance protein LytB.  
 GN LYTB OR SO3529.  
 OS Shewanella oneidensis.  
 OC Bacteria; Proteobacteria; Gammaproteobacteria; Alteromonadales;  
 OC Alteromonadaceae; Shewanella.  
 CX NCBI\_TaxID=70863;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=NR-1;  
 RX MEDLINE=22297686; PubMed=12366813;  
 RA Heidelberg J.E., Paulsen I.T., Nelson K.E., Gaidos E.J., Nelson W.C.,  
 Read T.D., Eisen J.A., Seshadri R., Ward N., Methe B., Clayton R.A.,  
 Meyer T., Tsapin A., Scott J., Beaman M., Brinkac L., Daugherty S.,  
 Debey R.T., Pederson R.J., Durkin A.S., Haft D.H., Kolonay J.F.,  
 Madupu R., Peterson J.D., Umayam L.A., White O., Wolf A.M.,  
 Vamathevan J., Weidman J., Imperial M., Lee K., Berry K., Lee C.,  
 Mueller J., Khouri H., Gill J., Usterback T.R., McDonald L.A.,  
 Feldblyum T.V., Smith H.O., Venter J.C., Nelson K.H., Fraser C.M.,  
 RT "Genome sequence of the dissimilatory metal ion-reducing bacterium  
 Shewanella oneidensis."  
 RL Nat. Biotechnol. 20:1118-1123 (2002).  
 DR EMBL: AE015789; AAN56520.1; -.  
 DR TIGR: SO3529; -.  
 KW Complete proteome.  
 SQ SEQUENCE 318 AA; 34718 MW; ADCE10B6A3582E82 CRC64;

Query Match 59.0%; Score 958; DB 16; Length 318;  
 Best Local Similarity 60.1%; Pred. No. 5,7e-64;  
 Matches 188; Conservative 50; Mismatches 71; Indels 4; Gaps 2;

QY 1 MQVLANPRGFCAGVDRAIEIVDAIEAFGAPIYVREHVHNRVYVDGLKQKAVFIEEL 60  
 DB 10 LNIKLANPRGFCAGVDRAISIVERALELFSPEIYVREHVHNRVYVDGLKQKAVFIEEL 69  
 QY 61 SDVPVGSYLIFSAHGVSKVQOAEEROLTVFATCPVATKVMQVAKHAKQGEVLIG 120  
 DB 70 DQVFNNTYVIFSAHGVSKVQVRAKARGLVFDAICPLVTKHDLVYTAASKGIECLIG 129  
 QY 121 HAHGEVEGTMGQYEKTEGGGIYLVETPEDVRNLKVNPNDLAVYVOTLLSMOTDKMV 180  
 DB 130 HAHGEVEGTMGQYD--NPNNGSVYLIESPADVETLEVDPNNLCTVGTTLISVDTLDI 187  
 QY 181 DALREQPSIKEOKKDDICATATONRODAVHDLAKISDLILVYSGSPSSNSNRLRELAIV 240  
 DB 188 SELLERFPIBGRKDDICATATONRODAVHDLAKISDLILVYSGSPSSNSNRLRELAIV 247  
 QY 241 GKPAVILIDYODIKODMLBGIEVGVTAGASAEVLVQVVDQLKAMGGETTSVRENSGI 300  
 DB 248 GTOSTLVITADIDISWENITKAVTAGASAEVLVQVVDQLKAMGGETTSVRENSGR 305  
 QY 301 BEKVFSPKPELK 313

DB 306 KEDTFAVPAELR 318

## RESULT 3

Q8D2R2 PRELIMINARY; PRT; 315 AA.  
 ID Q8D2R2  
 AC Q8D2R2  
 DT 01-MAR-2003 (Tremblrel. 23, Created)  
 DT 01-MAR-2003 (Tremblrel. 23, Last sequence update)  
 DT 01-MAR-2003 (Tremblrel. 23, Last annotation update)  
 DE LytB protein.  
 GN LYTB.  
 OS Myxlesworthia brevipalpis.  
 OC Bacteria; Proteobacteria; Gammaproteobacteria; Enterobacteriales;  
 OC Enterobacteriaceae; Myxlesworthia.  
 CX NCBI\_TaxID=164609;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC MEDLINE=22297718; PubMed=12219091;  
 RA Akman L., Yamashita A., Watanabe H., Oshima K., Shiba T., Hattori M.,  
 Aksoy S.,  
 RT "Genome sequence of the endocellular obligate symbiont of tsetse  
 flies, Myxlesworthia glosiniida."  
 RL Nat. Genet. 32:402-407 (2002).  
 DR EMBL: AB063521; BAC24438.1; -.  
 KW Complete proteome.  
 SQ SEQUENCE 315 AA; 35860 MW; 6552098712AA2884 CRC64;

Query Match 53.4%; Score 866; DB 16; Length 315;  
 Best Local Similarity 51.9%; Pred. No. 4,6e-57;  
 Matches 162; Conservative 72; Mismatches 76; Indels 2; Gaps 1;

QY 1 MQVLANPRGFCAGVDRAIEIVDAIEAFGAPIYVREHVHNRVYVDGLKQKAVFIEEL 60  
 DB 1 MNIPLANPRGFCAGVDRAIIVKRAIEIYGPPIYVREHVHNRVYVDGLKQKAVFIEEL 60  
 QY 61 SDVPVGSYLIFSAHGVSKVQOAEEROLTVFATCPVATKVMQVAKHAKQGEVLIG 120  
 DB 61 CDVPSKILIFSAHGVSKVSKLKEANKKLLIFATCPVSKVHHEIKGASKSEVLIIG 120  
 QY 121 HAHGEVEGTMGQYEKTEGGGIYLVETPEDVRNLKVNPNDLAVYVOTLLSMOTDKMV 180  
 DB 121 HXNHPIIGTIGQYNNPNK--KVFIQSTIEICLAKIKDNNLFFYGTTLISVDTLTKII 178  
 QY 181 DALREQPSIKEOKKDDICATATONRODAVHDLAKISDLILVYSGSPSSNSNRLRELAIV 240  
 DB 179 FAIKKKYPIIIBPKDDICATATONROKSIKKIKLIVDIIPIISKSSNSNRLRELAIV 238  
 QY 241 GKPAVILIDYODIKODMLBGIEVGVTAGASAEVLVQVVDQLKAMGGETTSVRENSGI 300  
 DB 239 GKSKSYLIDYKEIKKSLNGLNNGNIGITKASAEVLVQVVDQLKAMGGETTSVRENS 298  
 QY 301 BEKVFSPKPELK 312  
 DB 299 IETKFMIPKEL 310

## RESULT 4

Q8FQPO PRELIMINARY; PRT; 336 AA.  
 ID Q8FQPO  
 AC Q8FQPO  
 DT 01-MAR-2003 (Tremblrel. 23, Created)  
 DT 01-MAR-2003 (Tremblrel. 23, Last sequence update)  
 DT 01-MAR-2003 (Tremblrel. 23, Last annotation update)  
 DE Putative isoprenyl diphosphate biosynthesis protein Isph.  
 GN CEI079.  
 OS Corynebacterium efficiens.  
 OC Bacteria; Actinobacteria; Actinobacteridae; Actinomycetales;  
 OC Corynebacteriaceae; Corynebacteriaceae; Corynebacterium.  
 CX NCBI\_TaxID=152794;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=Y8-314 / AJ 12310 / DSM 44549 / JCM 11189;

RA Kawarabayasi Y., Yamazaki J., Hino Y., Kikuchi H., Nakamura Y.,  
 RA Ikeo K., Suzuki M., Mashima J., Itoh T., Yamagishi A., Nishio Y.,  
 RA Usuda Y., Sugimoto S.,  
 RT "The entire genomic sequence of *Corynebacterium efficiens* YS-314."  
 RT Submitted (MAY-2002) to the EMBL/GenBank/DBJ databases.  
 DR EMBL; AB005217; BAC17889.1; -.  
 KW Complete proteome.  
 SQ SEQUENCE 336 AA; 36430 MW; C47B456FEEB82AC7 CRC64;

Query Match 47.1%; Score 764; DB 16; Length 336;  
 Best Local Similarity 48.4%; Pred. No. 2,3e-49;  
 Matches 153; Conservative 68; Mismatches 83; Indels 12; Gaps 5;

QY 2 QIVLANPFGCAGVDRAIETVDQIAEAFAPIVYRHEVHNRITVVDGLKQKQAVFIEEL 61  
 DB 26 KILLAAPRGYAGVRAVETVERALEEYGAPIYVREIYVHNRVYVDTLAEKGVIFVEAS 85  
 QY 62 DVPVSYLIFSAHGVSKVQQAEROLVFPATCPVTKYHMQVAKAKQREYILIGH 121  
 DB 86 EAPBGAHMFSAHGVSPAKVEPAKALQAIAPACPLVTKVNEVRRPDQKQFILLFIGH 145  
 QY 122 AGHPVEVGTMGQYEKCTEGGIIYVETPEDEVRLK---VNNENDLAYVYQTLTSMITKV 178  
 DB 146 BGHEEVGTMGHSLDRT-----HIVDGLIESIPGLAFIADEN-LIYLSQTLTSDVETME 199  
 QY 179 MVDALREQPSIKQKDDICATQNRQDAVHDLAKISDLILVSGSPNSNSNRLEIAY 238  
 DB 200 IYRELKRYPPQLQDPSPSDICATQNRQDAVHDLAKISDLILVSGSPNSNSNRLEIAY 259  
 QY 239 QLG-PPAYLIDITVYQDKODLEGIEVGTAGASPEVLYOGVITDQKAMGETTSVEN 297  
 DB 260 QGAGRAHLYVDAHIDEAMLDGVCTIGISSASPELITVGLERLAGYGRD-DVSEV 317  
 QY 298 SGIEEKVVFISIPKELK 313  
 DB 318 TTAARKIVFALPRVLR 333

## RESULT 5

ID Q8G257 PRELIMINARY; PRT; 372 AA.  
 AC Q8G257;  
 DT 01-MAR-2003 (TREMblrel. 23, Created)  
 DT 01-MAR-2003 (TREMblrel. 23, Last sequence update)  
 DT 01-MAR-2003 (TREMblrel. 23, Last annotation update)  
 DE Penicillin tolerance protein lypB.  
 DE lypB OR BR0475.  
 OS Brucella suis.  
 OC Bacteria; Proteobacteria; Alphaproteobacteria; Rhizobiales;  
 OC Brucellaceae; Brucella.  
 OX NCBI\_TaxID=29461;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=1330 / Biovar 1;  
 RX MEDLINE=22247741; PubMed=12271122;  
 RA Paulsen I.T., Seshadri R., Nelson K.E., Eisen J.A., Heidelberg J.F.,  
 RA Read T.D., Dodson R.J., Umayam L., Brinkac L.M., Beaman M.J.,  
 RA Daugherty S.C., Debey R.T., Durkin A.S., Kolonay J.F., Madupu R.,  
 RA Nelson W.C., Ayodeji B., Kraul M., Sletty J., Malek U., Van Aken S.E.,  
 RA Rietmuller S., Tettelin H., Gill S.R., White O., Salzberg S.L.,  
 RA Hoover D.L., Lindler L.B., Hailing S.M., Boyle S.M., Fraser C.M.,  
 RT "The Brucella suis genome reveals fundamental similarities between  
 RT animal and plant pathogens and symbionts."  
 RT Proc. Natl. Acad. Sci. U.S.A. 99:13148-13153(2002).  
 DR EMBL; AE014357; AAN29418.1; -.  
 KW Complete proteome.  
 SQ SEQUENCE 372 AA; 40930 MW; 365B9906518CC9BC CRC64;

Query Match 46.3%; Score 751; DB 16; Length 372;  
 Best Local Similarity 47.7%; Pred. No. 2,6e-48;  
 Matches 154; Conservative 59; Mismatches 88; Indels 22; Gaps 5;

QY 1 MQIVLANPFGCAGVDRAIETVDQIAEAFAPIVYRHEVHNRITVVDGLKQKQAVFIEEL 60  
 DB 34 LEIRLCGRGRCAGVDRAIQIVLALCKKGAIPYVRRHIVNRRYVBEGLQARGAIFVEEL 93  
 QY 61 SDVPV---GSTYLFSAHGVSKVQQAEROLVFPATCPVTKYHMQVAKAKQREYI 117  
 DB 94 DBIPAAHNRQPVFSAHGVSPAKVEPAKALFYIDATCPVSKYKQAMHQRIGRAVI 153  
 QY 118 LIHAGHPVEVGTMGQYEKCTEGGIIYVETPEDEVRLK---VNNENDLAYVYQTLTSMITKV 177  
 DB 154 LIHAGHPVEVGTMGQ-----LPDQAVLITVEVDHATCHPDEDNLGTYQTTLTSDVDTA 209  
 QY 178 MVDALREQPSIKQKDDICATQNRQDAVHDLAKISDLILVSGSPNSNSNRLEIAY 237  
 DB 210 GIIEKLQARFPNLAAFAAESICATQNRQDAVRAAPQCDLFIYGAAPNSNSNRLEIAY 269  
 QY 238 VQLG-KPAYLIDITVYQDKODLEGIEVGTAGASPEVLYOGVITDQKAMGETTSVEN 290  
 DB 270 EKQGRASMLVQRAEDLEMEQIGDISVGLSAGASAPETIVDBITIDAKAFPDVKIELAE 329  
 QY 291 TTSVRENSGIEEKVVFISIPKELK 313  
 DB 330 TTVETEN-----FLVNRER 344

## RESULT 6

ID Q8G418 PRELIMINARY; PRT; 352 AA.  
 AC Q8G418;  
 DT 01-MAR-2003 (TREMblrel. 23, Created)  
 DT 01-MAR-2003 (TREMblrel. 23, Last sequence update)  
 DT 01-MAR-2003 (TREMblrel. 23, Last annotation update)  
 DE lppH protein involved in conversion of  
 DE 1-hydroxy-2-methyl-2-(E)-butenyl 4-diphosphate into lpp and  
 DE DMAP.  
 DE ISPH OR BL1361.  
 OS Bifidobacterium longum  
 OC Bacteria; Actinobacteria; Actinobacteridae; Bifidobacteriales;  
 OC Bifidobacteriaceae; Bifidobacterium.  
 OX NCBI\_TaxID=216816;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=NCC 2705;  
 RX MEDLINE=22294977; PubMed=12381787;  
 RA Schell M.A., Karimiantzou M., Snel B., Vilanova D., Berger B.,  
 RA Pessi G., Zwahlen M.-C., Desiere F., Bork P., Delley M.,  
 RA Pidmore R.D., Arigoni F.,  
 RT "The genome sequence of *Bifidobacterium longum* reflects its adaptation  
 RT to the human gastrointestinal tract."  
 RT Proc. Natl. Acad. Sci. U.S.A. 99:14422-14427(2002).  
 DR EMBL; AE014766; AAN25161.1; -.  
 KW Complete proteome.  
 SQ SEQUENCE 352 AA; 37863 MW; 984B6BF80286D136 CRC64;

Query Match 40.5%; Score 658; DB 16; Length 352;  
 Best Local Similarity 43.8%; Pred. No. 2,3e-41;  
 Matches 145; Conservative 64; Mismatches 96; Indels 26; Gaps 7;

QY 3 VIANPFGCAGVDRAIETVDQIAEAFAPIVYRHEVHNRITVVDGLKQ 51  
 DB 27 VVADLPFGCAGVDRAIETVDQIAEAFAPIVYRHEVHNRITVVDGLKQ 86  
 QY 52 KGAVFIEELSDVP-----VGSYLFSAHGVSKVQQAEROLVFPATCPVTKYHMQV 106  
 DB 87 QGAVFVQELAEIPDAQAAGIPIVFSHGVSPVKAERGMHVDATCPVTKYHMQV 146  
 QY 107 AKAKQREYILIGHAGHPVEVGTMGQYEKCTEGGIIYVETPEDEVRLK---VNNENDLAYV 166  
 DB 147 LRFVARGYRIVYIGHKGDEAVGVGSP-----HVHLIEBSDDVDSLPADTKLVIL 201  
 QY 167 TQTLTSMITKVAVDALARQPSIKQKDDICATQNRQDAVHDLAKISDLILVSGSPN 226  
 DB 202 SQTLTSDVETADITALKAKFPWIOEPSSSDICATQNRQDAVHDLAKISDLILVSGSAN 261

QY 227 SSNSNRRLREIAVO-LGK--PAYLIDTYODLKQDMLEGEVGVTAASAEVLYVOEVIDQ 283  
 DB 262 SSNSNRRLREIAVO-LGK--PAYLIDTYODLKQDMLEGEVGVTAASAEVLYVOEVIDQ 321  
 QY 284 IKAMGERTSVRENSGIEEKVVFSPKELK 314  
 DB 322 LONLQ--FTGMRKSVETIKENHFFVLPALRLR 350

## RESULT 7

Q88313 PRELIMINARY; PRT; 330 AA.

AC Q88313; 01-MAR-2003 (TREMBlrel. 23, Created)  
 DT 01-MAR-2003 (TREMBlrel. 23, Last sequence update)  
 DE 01-MAR-2003 (TREMBlrel. 23, Last annotation update)  
 GN Penicillin tolerance protein.  
 OS Leptospira interrogans.  
 OC Bacteria; Spirochaetes; Spirochaetales; Leptospiraceae; Leptospira.  
 RX NCBI\_TaxID=173;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=56601 / Serogroup Icterohaemorrhagiae / Serovar 1a1;  
 RA Ren S.  
 RT Submitted (MAR-2002) to the EMBL/GenBank/DBJ databases.  
 DR EMBL; AE011409; AAK49619.1;  
 KW Complete proteome.  
 SQ SEQUENCE 330 AA; 36940 MW; 5D7642CD6AE3F86E CRC64;

Query Match 38.7%; Score 628; DB 16; Length 330;  
 Best Local Similarity 47.1%; Pred. No. 3.8e-39;  
 Matches 136; Conservative 57; Mismatches 82; Indels 14; Gaps 7;

QY 2 QVLANPFGCGVDRALREIYDQALFAFGAPRYVREHVNRVTVDGKQKGAFTIEHSD 61  
 DB 22 KYLANPFGCGVDRALREIYDQALFAFGAPRYVREHVNRVTVDGKQKGAFTIEHSD 81  
 QY 62 DVPVGSYLIFSAHGVSKVQOEAEEROLTVFDPATCELYTKVMQVAKAKQREYILIGH 121  
 DB 82 EAVNKGATVIFSAHGVSPVVEAKQKGMKIGDAPCELYTVRHRK-ARKIKDTHQIYIGH 140  
 QY 122 AGHPVEVGMQYKCTBGGGIYLVETPEDVRLKKN-NPN-DLAVTQTTLSMTDTK 179  
 DB 141 EGHDEAIGTGEAE-----MFLVESLDDIISLKDIPNKRLTYLQTTLSVADTK 193  
 QY 180 VDALRQFPSTIKQKQKQDICYATONRODAVHDLAKISDLILVGSPPSSNSNRRLREIAVO 239  
 DB 194 IDQISTFFVEHPSKDDICVATTEKQEAVALMDKIDMLVIGADNSNSRLRLQLA-Q 252  
 QY 240 LGKP-AYLIDTYODLKQDMLEGEVGVTAASAEVLYVOEVIDQ 285  
 DB 253 KSKPESHFKVSTADLSKEYIYONNEIKILGLITAGASTPYLVDEIISKLK 301

## RESULT 8

Q81295 PRELIMINARY; PRT; 535 AA.

AC Q81295; 01-MAR-2003 (TREMBlrel. 23, Created)  
 DT 01-MAR-2003 (TREMBlrel. 23, Last sequence update)  
 DE 01-MAR-2003 (TREMBlrel. 23, Last annotation update)  
 GN LytB protein.  
 OS Plasmodium falciparum (isolate 3D7).  
 OC Eukaryota; Alveolata; Apicomplexa; Haemosporida; Plasmodium.  
 RX NCBI\_TaxID=36339;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC MEDLINE=22255708; PubMed=12368867;  
 RA Hall N., Pain A., Bettman M., Churcher C., Harris B., Harris D., Mungall K., Bowman S., Atkin R., Baker S., Barron A., Brooks K.,

RA Buckee C.O., Burrows C., Cherevach I., Chillingworth C.,  
 RA Chillingworth T., Christodoulou Z., Clark L., Clark R., Corron C.,  
 RA Cronin A., Davies R., Davis P., Dear P., Dearden F., Doggett J.,  
 RA Fellwell T., Goble A., Goodhead I., Gwilliam R., Hamlin N., Hance Z.,  
 RA Harper D., Hauser H., Hornsby T., Holroyd S., Horrocks P.,  
 RA Humphrey S., Jagels K., James K.D., Johnson D., Kerhornum A.,  
 RA Knights A., Konfortov B., Kyes S., Larke N., Lawson D., Lennard N.,  
 RA Line A., Maddison M., McLean J., Mooney P., Moule S., Murphy L.,  
 RA Oliver K., Ormond D., Price C., Quail M.A., Rabinowitsch E.,  
 RA Rajandream M.A., Rutter S., Rutherford K.M., Sanders M., Simmonds M.,  
 RA Seeger K., Sharp S., Smith R., Squares R., Squares S., Stevens K.,  
 RA Taylor K., Tivey A., Unwin L., Whitehead S., Woodward J.,  
 RA Tulston J.E., Craig A., Newbold C., Barrett B.G.  
 RT "Sequence of Plasmodium falciparum chromosomes 1, 3-9 and 13."  
 RL Nature 419:527-531(2002).  
 DR EMBL; AL011745; CAD99005.1;  
 SQ SEQUENCE 535 AA; 62470 MW; E7CB6FCCD22FA07E CRC64;

Query Match 36.9%; Score 599; DB 5; Length 535;  
 Best Local Similarity 41.4%; Pred. No. 1.1e-36;  
 Matches 130; Conservative 58; Mismatches 120; Indels 6; Gaps 2;

QY 3 IVLANPFGCGVDRALREIYDQALFAFGAPRYVREHVNRVTVDGKQKGAFTIEHSD 62  
 DB 222 KYLANPFGCGVDRALREIYDQALFAFGAPRYVREHVNRVTVDGKQKGAFTIEHSD 281  
 QY 63 DVPVGSYLIFSAHGVSKVQOEAEEROLTVFDPATCELYTKVMQVAKAKQREYILIGH 122  
 DB 282 VPDHHLILYSAHGISPQREIAKKKKLEIDATCELYTKVMQVAKAKQREYILIGH 341  
 QY 123 GHPVEVGMQYKCTBGGGIYLVETPEDVRLKKN-NPN-DLAVTQTTLSMTDTK 182  
 DB 342 NHVEVIGTGNAPRCT-----HIVENVDVKNLFPNKKLFYVQTTLSMDALYQK 396  
 QY 183 LRQFPSTIKQKQKQDICYATONRODAVHDLAKISDLILVGSPPSSNSNRRLREIAVO 242  
 DB 397 LRKFPSTIKQKQKQDICYATONRODAVHDLAKISDLILVGSPPSSNSNRRLREIAVO 456  
 QY 243 PAYLIDTYODLKQDMLEGEVGVTAASAEVLYVOEVIDQ 285  
 DB 457 PAYLIDTYODLKQDMLEGEVGVTAASAEVLYVOEVIDQ 301  
 QY 303 KVPSPKELK 316  
 DB 517 VPKMKLPNML-LHM 529

## RESULT 9

Q9BX6 PRELIMINARY; PRT; 535 AA.

AC Q9BX6; 01-JUN-2001 (TREMBlrel. 17, Created)  
 DT 01-JUN-2001 (TREMBlrel. 17, Last sequence update)  
 DE 01-MAR-2003 (TREMBlrel. 23, Last annotation update)  
 GN LytB.  
 OS Plasmodium falciparum.  
 OC Eukaryota; Alveolata; Apicomplexa; Haemosporida; Plasmodium.  
 RX NCBI\_TaxID=5833;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC MEDLINE=21172655; PubMed=11274098;  
 RA Alincic B., Kollas A.K., Sanderbrand S., Wiesner J., Hintz M.,  
 RA Beck E., Juma H.,  
 RT "Gp6 is involved in the 2-C-methyl-D-Erythritol 4-Phosphate pathway  
 of Isoprenoid Biosynthesis in Escherichia coli."  
 RL J. Bacteriol. 183:2411-2416(2001).  
 DR EMBL; AF333927; AAK12102.1;  
 DR InterPro; IPR003451; LytB.  
 DR InterPro; IPR002052; N6\_Mtase.  
 DR Pfam; PF02401; LytB; 1.  
 DR TIGRFAMs; TIGR00216; 1spH\_lytB; 1.  
 DR PROSITE; PS00092; N6\_Mtase; 1.

SEQUENCE 535 AA; 62504 MW; E7C6DCDD22FA07E CRC64;  
 Query Match 36.9%; Score 598.5; DB 5; Length 535;  
 Best Local Similarity 43.6%; Pred. No. 1,3e-36;  
 Matches 123; Conservative 55; Mismatches 99; Indels 5; Gaps 1;  
 3 IYANRGRGCGAGVRAIEIVDAIEAFGAPIYRHEVNRKTVVDGLKQKGAFFIEBLSD 62  
 222 LVLINRGRGCGAGVRAIEIVDAIEAFGAPIYRHEVNRKTVVDGLKQKGAFFIEBLSD 281  
 63 VAVGSLIFASAGVSEVNOQAEERQITVDPATCPLVTVNQVAKHAKQSEVILLIGA 122  
 282 VDPGHLITVSAIGISQPIREIAKCKLIEIDTCPLVNVVHYVQKAKENYDILLIGK 341  
 123 GHPEVGTGQYKTEGEGGIVLVEPDEVRLKVNPNDAVYVQTLSMTPTKVMVA 182  
 342 NVEVIGITNEAPHQ-----HIVEVNDVVDKLNPFINKCLPVYVQTLSMDCLIVQK 396  
 183 LREQPSIKEQKDDICVATONRODAVHDLAKISDILLVSSPNSNSRLREIVQJGK 242  
 397 LKXKPEHIEIIPSGSICVATTKQTMINKICTKCDLTIIVGSSSSNAKCLVYSSQIRVY 456  
 243 PAVLIDTVQDLKQDMLEGIEVVGVTAGASAPVLYVQEVLDL 284  
 457 PAVLIDTVQDLKQDMLEGIEVVGVTAGASAPVLYVQEVLDL 498

## RESULT 10

OBERS12 PRELIMINARY; PRT; 827 AA.  
 AC OBERS12; (TREMELrel. 21, Created)  
 DT 01-JUN-2002 (TREMELrel. 21, Last sequence update)  
 DT 01-MAR-2003 (TREMELrel. 23, Last annotation update)  
 DE LysB protein.  
 GN FN1781.  
 OS Fusobacterium nucleatum (subsp. nucleatum).  
 OC Bacteria; Fusobacteriia; Fusobacteriales; Fusobacteriaceae;  
 CC Fusobacterium.  
 OX NCBI\_TaxID=76856;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=ATCC 25586;  
 RX MEDLINE=2186394; PubMed=11889109;  
 RA Kapelral V., Anderson I., Ivanova N., Reznik G., Los T., Lykidis A.,  
 RA Bhattacharya A., Barton A., Gardner W., Grechlin G., Zhu L.,  
 RA Vasileva O., Chu L., Kogan Y., Chaga O., Goltsman E., Bernal A.,  
 RA Larsen N., D'Souza M., Malinas T., Pusch G., Haselkorn R.,  
 RA Fongstein M., Kyriades N., Overbeek R.;  
 RT Genome sequence and analysis of the oral bacterium Fusobacterium  
 RT nucleatum strain ATCC 25586.  
 RL J. Bacteriol. 184:2005-2018(2002).  
 DR EMBL; AE010480; AAL93880.1; -  
 DR InterPro; IPR003451; LysB.  
 DR InterPro; IPR003029; SL.  
 DR Pfam; PF02401; LysB; 1.  
 DR Pfam; PF00575; SL; 4.  
 DR SMART; SM00316; SL; 6.  
 DR TIGRPFAM; TIGR00216; lspb\_LysB; 1.  
 DR PROSITE; PS0126; SL; 4.  
 DR Complete proteome.  
 SQ SEQUENCE 827 AA; 94507 MW; 23871DGED082A35E CRC64;

Query Match 24.9%; Score 404.5; DB 16; Length 827;  
 Best Local Similarity 30.7%; Pred. No. 8.9e-22;  
 Matches 103; Conservative 68; Mismatches 119; Indels 45; Gaps 9;  
 1 MOIVLANRGRGCGAGVRAIEIVDAIEAFGAPIYRHEVNRKTVVDGLKQKGAFFIEB 59  
 1 MEIIRAKHMGFCGVRAIIVNCSLYEGRK-YIIIGMLVHKKQVDEMERGKFLVIED 59  
 60 -----LSIDVPGSYLIFSAGVSEVNOQAEERQITVDPATCPLVTVNQVAKHAKQ 114

DB 60 ELLNMDLEKEDDIIVIRAHGTSKVEHKLKRVKVPDTCIPVNRKIRQELIANENGY 119  
 QY 115 EVIILIGHGHEVBTMQYKTEGEGGIVLVEPDEVRLKVNPNDAVYVQTLSMT 174  
 DB 120 SILFEGDNHPEVQV-----SFADIQIFESFEKAKKIDIDKTYLLSTQTLN-- 171  
 QY 175 DTKVVDALREQPSIKEQK-----DDICVATONRODAVHDLAKISDILLVWSP 225  
 DB 172 -----KKKFEIKKFKFENKVNVIIPKICATVROKAVEDLVANVEMIVYD 222  
 QY 226 MSSNSRLREIVQJGKPAVYIDTVQDLKQDMLEGIEVVGVTAGASAPVLYVQEVLD 285  
 DB 223 KSNNTKLYEISKTLNDSYLVNEBQDLSTIPKGEVVGITAGASTPEETIMNI--EK 280  
 QY 286 AMG-GETTSVRENSG-----TEEKVPSIPKELK 314  
 DB 281 VRIYKMSNVNENQERSLMEB-----FLPNQERK 311

## RESULT 11

OBERS9 PRELIMINARY; PRT; 299 AA.  
 AC OBERS9; (TREMELrel. 23, Created)  
 DT 01-MAR-2003 (TREMELrel. 23, Last sequence update)  
 DT 01-MAR-2003 (TREMELrel. 23, Last annotation update)  
 DE LysB protein.  
 GN MYEL1330.  
 OS Mycoplasma penetrans.  
 OC Bacteria; Firmicutes; Mollicutes; Mycoplasmataceae; Mycoplasma.  
 OX NCBI\_TaxID=28227;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=HF-2;  
 RX MEDLINE=22354719; PubMed=1246655;  
 RA Saeki Y., Ishikawa J., Yamashita A., Oshina K., Kenti T., Furuya K.,  
 RA Yoshino C., Horino A., Shida T., Saeki T., Hattori M.;  
 RT The complete genomic sequence of Mycoplasma penetrans, an  
 RT intracellular bacterial pathogen in humans.  
 RL Nucleic Acids Res. 30:5293-5300(2002).  
 DR EMBL; AP004170; BAC43925.1; -  
 DR Complete proteome.  
 SQ SEQUENCE 299 AA; 34190 MW; CFF14C65FB55F9E CRC64;

Query Match 24.2%; Score 393; DB 16; Length 299;  
 Best Local Similarity 30.2%; Pred. No. 1.5e-21;  
 Matches 93; Conservative 68; Mismatches 99; Indels 48; Gaps 8;

QY 1 MOIVLANRGRGCGAGVRAIEIVDAIEAFG-----APIYRHE-----VYHNRVVDGL 49  
 DB 1 MKIITKTPRGFCGV-----VDAYATCKKIATKLYPNHEKYLIGLVNHEIKEL 50  
 QY 50 KQK-----AVFTELSDVPGSYLIFSAGVSEVNOQAEERQITVDP 93  
 DB 51 EELGIQTMDDKNSRSEIIDSIEIKDKNRP-----VIFSAGTQKTDKAREKGLVFD 107  
 QY 94 ATCPVTVYVQVAKHAKQSEVILLIGHAGPEVGTMGQYKTEGEGGIVLVEPDEV 153  
 DB 108 TTCIYVPTKDLIRKEKEQGYQIFYGNNHPEITSL-SIDK-----SILLIEVNDIB 161  
 QY 154 NLKVNPNDAVYVQTLSMTDKVMDALREQPSIKEQKDDICVATONRODAVHDL 213  
 DB 162 NRTESKEPIVYVQTLSMTSIFEFIEITELSKRNI--EKIDICNAKRODAVINMP 219  
 QY 214 KISDILLVSSPNSNSRLREIVQJGKPAVYIDTVQDLKQDMLEGIEVVGVTAGASAP 273  
 DB 220 SEVLLVAVGDIKNSNKKLVIGIKQIESHLIMFNKNIKDEWFINKKCLATISGCTP 279  
 QY 274 EVLYVQEV 281  
 DB 280 TWLANVY 287

RESULT 12  
 097109 PRELIMINARY: PRT; 642 AA.  
 AC 097109: 18, Created)  
 DT 01-OCT-2001 (TREMELrel. 18, last sequence update)  
 DT 01-MAR-2003 (TREMELrel. 23, last annotation update)  
 DE Fusion penicillin tolerance lytB domain (N-term) and S1 ribosomal protein (C-term).  
 GN CAC1847.  
 OS Clostridium acetobutylicum.  
 OC Bacteria; Firmicutes; Clostridia; Clostridiales; Clostridiaceae;  
 OC Clostridium.  
 OX NCBI\_TaxID=1488;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=ATCC 824 / DSM 792 / VKM B-1787;  
 RX MEDLINE=21359325; PubMed=11466286;  
 RA Noelling J., Breton G., Ometchenko M.V., Makarova K.S., Zeng Q.,  
 RA Gibson R., Lee H.M., Dubois J., Qiu D., Hilt J., Wolf Y.I.,  
 RA Tatusov R.L., Sabathe F., Doucet-Stamm L., Soucaille P., Daly M.J.,  
 RA Bennett G.N., Koonin E.V., Smith D.R.;  
 RA Genome sequence and comparative analysis of the solvent-producing  
 RT bacterium Clostridium acetobutylicum.  
 RL J. Bacteriol. 183:4823-4838(2001).  
 DR EMBL; AF07693; AAK79811.1; -  
 DR InterPro; IPR003451; LytB.  
 DR InterPro; IPR000110; Ribosomal\_S1.  
 DR InterPro; IPR003029; S1.  
 DR Pfam; PF02401; LytB; 1.  
 DR Pfam; PF0575; S1; 3.  
 DR PRINTS; PR00681; RIBOSOMAL\_S1.  
 DR SMART; SM00316; S1; 4.  
 DR TIGRFAMs; TIGR00216; lspb\_lytB; 1.  
 DR PROSITE; PS0126; S1; 3.  
 DR Complete proteome.  
 KW Complete proteome.  
 SQ SEQUENCE 642 AA; 71968 MW; A8833AA24420EBD CRC64;

Query Match 22.1%; Score 358; DB 16; Length 642;  
 Best Local Similarity 27.1%; Pred. No. 1.9e-18;  
 Matches 90; Conservative 77; Mismatches 129; Indels 36; Gaps 8;

097109 PRELIMINARY: PRT; 402 AA.  
 AC 097109: 23, Created)  
 DT 01-MAR-2003 (TREMELrel. 23, last sequence update)  
 DT 01-MAR-2003 (TREMELrel. 23, last sequence update)

DT 01-MAR-2003 (TREMELrel. 23, last annotation update)  
 DE TIR1041 protein.  
 GN TIR1041.  
 OS Synecococcus elongatus (Thermosynechococcus elongatus).  
 OC Bacteria; Cyanobacteria; Chroococcales; Synecococcus.  
 OX NCBI\_TaxID=32046;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=BP-1;  
 RX MEDLINE=2225144; PubMed=12240834;  
 RA Nakamura Y., Kaneko T., Sato S., Ikeuchi M., Katoh H., Sasamoto S.,  
 RA Watanabe A., Iriyuchi M., Kawashima K., Kimura T., Kishida Y.,  
 RA Kiyokawa C., Kohara M., Matsumoto M., Matsuno A., Nakazaki N.,  
 RA Shimizu S., Sugimoto M., Takeuchi C., Yamada M., Tabata S.;  
 RA Complete genome structure of the thermophilic cyanobacterium  
 RT Thermosynechococcus elongatus BP-1.  
 RL DNA Res. 9:123-130(2002).  
 DR EMBL; AF005372; BAC08594.1; -  
 DR Complete proteome.  
 KW Complete proteome.  
 SQ SEQUENCE 402 AA; 45213 MW; FB0851B61873510 CRC64;

Query Match 21.1%; Score 343; DB 16; Length 402;  
 Best Local Similarity 30.7%; Pred. No. 1.3e-17;  
 Matches 109; Conservative 53; Mismatches 105; Indels 88; Gaps 14;

097109 PRELIMINARY: PRT; 464 AA.  
 AC 097109: 16, Created)  
 DT 01-MAR-2001 (TREMELrel. 16, last sequence update)  
 DT 01-MAR-2001 (TREMELrel. 16, last sequence update)  
 DE LytB-like protein precursor.  
 GN LytB.  
 OS Adonis palaeolina.  
 OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;  
 OC Spermatophyta; Magnoliophyta; eudicotyledons; Ranunculales;  
 OC Ranunculaceae; Adonis.  
 OX NCBI\_TaxID=151078;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC Tissue=Immature and developing flower buds;  
 RX MEDLINE=20461233; PubMed=11004185;  
 RA Cunningham P.X., Jr., Lafond T.P., Gantt B.;  
 RA "Evidence of a Role for LytB in the Nonmevalonate Pathway of  
 RT Isoprenoid Biosynthesis."  
 RL J. Bacteriol. 182:5841-5848(2000).  
 DR EMBL; AF270978; AAG21984.1; -



